CYTOPLASMIC CIRCULAR

STABLE INTRONIC SEQUENCE RNA

by

Gaëlle Talhouarne

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

November, 2017

ABSTRACT

Introns represent one of the largest fractions of our genomes but also one of the least understood. Soon after transcription, most intronic RNAs are released as lariats and degraded within minutes; therefore, they are seen as discarded byproducts of splicing or junk. However, the Gall laboratory reported the existence of thousands of stable intronic sequence RNAs or sisRNAs in the oocyte nucleus (germinal vesicle) of the frog *Xenopus*. Also, highlighted in this thesis, I observed sisRNAs in the cytoplasm of these oocytes and eventually in other cells and species.

My thesis addresses the characterization of sisRNAs in the cytoplasm of the frog oocyte. I demonstrate that these transcripts are resistant to the exonuclease RNase R, and I confirm that most cytoplasmic sisRNAs accumulate as short lariats. I propose that sisRNAs can accumulate in the cytoplasm because this compartment lacks the lariat degradation machinery.

Additionally, I report that circular sisRNAs occur in human, mouse, chicken, and zebrafish as well. Most of these lariats are C-branched (unlike in *Xenopus*) and many accumulate in the cytoplasm. Surprisingly, hundreds of circular sisRNAs occur in human and mouse red blood cells, which lack nuclei and both transcription and translation machineries.

Finally, I catalog the circular sisRNAs which contain small nucleolar (sno)RNA motifs. These stable lariats bearing a snoRNA, or slb-snoRNAs, associate with the canonical snoRNA binding proteins, namely dyskerin and 15.5K, but do not function as canonical modification guide RNAs. Moreover, slb-snoRNAs can be exported to the

cytoplasm where they sequester dyskerin and 15.5K. I propose that other sisRNAs may have a regulatory role by sequestering other proteins.

Altogether, my studies strongly suggest that intronic RNAs are more than "junk" transcripts. Instead, they are selectively exported to the cytoplasm where they persist for hours. I will discuss their potential regulatory role.

Gaëlle J.S. Talhouarne

Joseph G. Gall, Ph.D. (advisor/reader)

Karen Beemon, Ph.D. (reader)

ACKNOWLEDGEMENT

First, I would like to express my sincere gratitude to my advisor, Dr. Joseph G. Gall, for the continuous support and guidance throughout my graduate career. His patience and love for science has been an invaluable inspiration. I could not have imagined having a better advisor and mentor for my Ph.D. study.

I would like to thank the rest of my thesis committee members Dr. Karen Beemon, Dr. Sarah Woodson, and Dr. Allan Spradling for their insight and encouragement.

For their scientific guidance and for providing a fun work environment, thank you to my labmates and lab meeting attendees, particularly to Svetlana Deryusheva, my colleague, collaborator, and friend.

I thank the faculty and staff of The Johns Hopkins University, Department of Biology, and the Carnegie Institution for Science, Department of Embryology.

Last but not least, a very special thanks to my husband as well as my parents, siblings, and aunt for their support.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Introns represent 20% of our genomes, yet they are poorly understood and even considered "junk." These non-coding (nc) sequences interrupt the coding regions of most protein-coding genes. Once transcribed, the intronic sequences are spliced from the pre-mRNA, released as lariats and degraded rapidly. The degradation pathway requires the RNA lariat debranching enzyme, Dbr1, which hydrolyzes the 2′–5′ covalent bond generated during splicing (Ruskin and Green 1985; Chapman and Boeke 1991; Ooi *et al*., 2001). Once linearized, intronic fragments are degraded by exonucleases in the nucleus (Hilleren and Parker, 2003).

Because introns are such an important energetic burden to cells, they have been the center of much argument regarding their origins and their functional capacity, debates for which little resolution has been offered. Many genomic comparative studies were carried out to establish the origin of introns and undertake the debate between the 'introns-early' (Doolittle, 1978; Darnell, 1978) and 'introns-late' (Doolittle and Stoltzfus, 1993; Mattick, 1994) hypotheses, asking whether introns are indigenous to ancestral genes or an invasion within eukaryotic genomes (Most recent comparative studies: Mourier and Jaffares, 2003; de Souza, 2003; Sverdlov *et al.*, 2003; Csűrös, 2005; Sverdlov *et al*., 2005; Jeffares *et al*., 2006; Li *et al*., 2009; Verhelst *et al*., 2013; Chen *et al*., 2017; Wu *et al*., 2017). The 'introns-early' and 'introns-late' examination shapes our understanding of intron function, specifically leading us to consider whether introns are gateways to protein evolution (exon shuffling) or selfish DNA sequences like transposons. Regardless, the percentage of intronic sequences increases with the 'complexity' of the eukaryotic species. Along with this observation, John Mattick (1994)

speculated that introns are more than selfish DNAs and instead may carry functional sequences for eukaryotic cells. We now know that certain introns contain noncoding RNAs, such as long non-coding (lnc) RNA, small nucleolar (sno) RNAs, or micro (mi) RNAs (Rearick *et al.*, 2011; Curtis *et al.*, 2012; Yin *et al.*, 2012), ncRNAs that regulate RNA metabolism. This list of encoded information within introns is only at its starting point.

Within the past few years, we have challenged the notion that other intronic RNAs are simply byproducts of splicing. Using high-throughput sequencing technology, the Gall laboratory analyzed the RNA from the giant oocyte nucleus or germinal vesicle (GV) of the frog *Xenopus tropicalis* (Gardner *et al.*, 2012). They expected to explore the nascent transcripts and, instead, found many intronic sequences in the GV. They proposed that these intronic sequences originated by splicing from pre-mRNA molecules. Because these sequences are extremely stable, we refer to them as stable intronic sequence RNA or sisRNA. Subsequently, by developing new tools, I discovered that a fraction of sisRNAs occurred also in the cytoplasm of amphibian oocytes. These observations are evidence that intronic RNAs are more than ncRNA hosts and I will argue in my thesis that intronic RNAs are functional ncRNAs.

The Discovery of cytoplasmic sisRNA

When I joined the Gall lab, I was interested in the cytoplasmic fraction of frog oocytes. Unlike most systems, the oocyte makes it possible to collect an extremely pure cytoplasmic fraction because the *intact* GV can be removed manually. In other words, I had a unique opportunity to search for ncRNAs occurring in the cytoplasm.

I began by analyzing high-throughput sequence data of non-ribosomal cytoplasmic RNA extracted from oocytes in which transcription was inhibited for 12 hrs. Most reads, *but not all*, mapped to exonic sequences. A small fraction of the reads mapped to introns instead. I developed new bioinformatic tools, with the help of Nicholas Ingolia, and identified the individual introns from which the intronic reads derived. This analysis generated the first evidence that sisRNAs accumulated in the cytoplasm. This observation raised three questions: What are the molecular characteristics of these intronic RNAs? Are cytoplasmic sisRNAs restricted to the *Xenopus* oocyte? What RNA binding proteins can associate with cytoplasmic sisRNA?

Molecular characteristics (Chapter 1)

In the GV, multiple sisRNAs could be produced from a single intron; therefore, we had assumed that a typical sisRNA was linear. With that in mind, I attempted to characterize the ends of several cytoplasmic sisRNA candidates, but failed. We now know that most if not *all* cytoplasmic sisRNAs are circular. They represent lariats that escaped the debranching enzyme. I took advantage of the fact the exonuclease RNase R degrades linear transcripts (including mRNAs) but not circular molecules to enrich for stable lariats and found about 9000 short cytoplasmic lariats.

At the time, very few studies employed RNase R; now, this enzyme is widely used to detect circRNA (exonic circular transcripts) in a variety of tissues and species (reviewed by Chen LL, 2016). Despite all the RNase R experiments published, for reasons I describe in the discussion, only two studies reported the existence of stable lariats (Zhang *et al.*, 2013 and Li *et al*., 2016). Nevertheless, I was encouraged when I

found lariats in multiple published high-throughput experiments and decided to explore further the possibility that sisRNAs occur in many species.

Circular sisRNA in other species (Chapter 2)

In large oocytes, only stable transcripts are detected by RNAseq (Gardner *et al.*, 2012), whereas, with a similar method, pre-mRNA and unstable lariats are detected in cultured cells or tissues (Taggart *et al.*, 2017). The contamination with such transient transcripts increases the difficulty in finding sisRNAs. Therefore, I began my search for stable lariats in cultured cells in which transcription was inhibited. I found hundreds of intronic RNAs resistant to RNase R in cultured human, mouse, chicken and frog cells.

As I mentioned earlier, the existence of cytoplasmic sisRNAs could be established in the first place only because I had analyzed a *pure* cytoplasmic fraction. Such efficient cell fractionation is difficult or impossible with most somatic cells. So instead, to prevent any contamination, I focused on red blood cells (RBCs). Circulating mammalian RBCs lack nuclei and thus provide a pure cytoplasm fraction. I detected 600 lariats in mouse RBCs and 300 in human. In chapter 2, I report further evidence for the existence of cytoplasmic sisRNAs in different vertebrate species.

RNA binding proteins and sisRNAs (Chapter 3)

The function of a ncRNA is often dictated by its RNA binding proteins (RBP). Hence, I searched for what RBPs bound sisRNAs. I used the capture hybridization analysis of RNA targets (CHART) method (Simon *et al.*, 2011) and pulled down the most abundant *Xenopus tropicalis* sisRNA, faf2, in oocyte lysate. Although the affinity

purification of sisRNA-faf2 was successful, too many unspecific proteins precipitated in both the sisRNA-faf2 pull down and control. I presume that the abundance of yolk proteins may be the cause of the high background. Alternatively, I took a candidate protein approach.

I noticed that some sisRNAs contained a snoRNA motif. Many snoRNAs guide rRNA and snRNA modifications when bound to the snoRNP proteins, including dyskerin (DKC1), an enzyme responsible for isomerization of uridine. In Chapter 3, I present a series of experiments suggesting that DKC1 can bind certain sisRNAs. In collaboration with Svetlana Deryusheva, we analyzed potential functions of such complexes.

7SL RNA in mature red blood cells (Chapter 4)

While analyzing mouse and human RBC data, I noticed that many other ncRNAs were readily detected, in addition to hundreds of lariats. A quarter to a third of reads mapped to the RN7SL gene, which encodes 7SL RNA or SRP RNA. Because this ncRNA facilitates the translation of transmembrane and secreted proteins, it was puzzling to detect 7SL RNA in cells lacking the translation machinery. In Chapter 4, I describe two further unexpected findings: mammalian RBCs contain a shorter ncRNA derived from the RN7SL gene, and 7SL RNA is associated with SART3, a non-canonical 7SL binding protein.

Altogether, I discovered a new face to intronic RNA. My studies shake preconceived notions regarding intronic RNA *and* 7SL RNA, while bringing to light a new layer of complexity. At the end of my thesis, I discuss the implications of such ncRNAs in unexpected places.

REFERENCES

Chapman KB, Boeke JD. **1991**. Isolation and characterization of the gene encoding yeast debranching enzyme. Cell 65: 483-492.

Chen B, Shao J, Zhuang H and Wen J. 2017. Evolutionary dynamics of triosephosphate isomerase gene intron location pattern in Metazoa: A new perspective on intron evolution in animals. *Gene*, *602*, pp.24-32.

Chen LL. **2016.** The biogenesis and emerging roles of circular RNAs**.** *Nature Reviews Molecular Cell Biology.* **17**, 205–211.

Csűrös, M., 2005. Likely scenarios of intron evolution. *Comparative genomics*, pp.47-60.

Curtis HJ, Sibley CR, Wood MJ. **2012**. Mirtrons, an emerging class of atypical miRNA. Wiley Interdiscip Rev RNA 3: 617-632.

Darnell JE Jr. **1978**. Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science. 202(4374):1257-60.

de Souza, S.J., 2003. The emergence of a synthetic theory of intron evolution. *Genetica*, *118*(2-3), pp.117-121.

Doolittle WF. **1978**. Genes in pieces: Were they ever together? Nature. 272:581–582.

Doolittle WF, Stoltzfus A. **1993**. Molecular evolution. Genes-in-pieces revisited. Nature. 361:403.

Gardner EJ, Nizami ZF, Talbot CC, Jr., Gall JG. **2012**. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of Xenopus tropicalis. Genes Dev 26: 2550-2559.

Hilleren PJ, Parker R. **2003**. Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates. Mol Cell 12: 1453-1465.

Jeffares, D.C., Mourier, T. and Penny, D., 2006. The biology of intron gain and loss. *TRENDS in Genetics*, *22*(1), pp.16-22.

Li, W., Tucker, A.E., Sung, W., Thomas, W.K. and Lynch, M., 2009. Extensive, recent intron gains in Daphnia populations. *Science*, *326*(5957), pp.1260-1262.

Li Z, Wang S, Cheng J, Su C, Zhong S, Liu Q, Yuda Fang Y, Yu Y, Lv H, Zheng Y and Zheng B. **2016**. Intron Lariat RNA Inhibits MicroRNA Biogenesis by Sequestering the Dicing Complex in *Arabidopsis.* PLoS Genet. 12(11): e1006422.

Mattick JS. **1994**. Introns: evolution and function. Curr Opin Genet Dev. 4:823–831.

Mourier, T. and Jeffares, D.C., **2003**. Eukaryotic intron loss. *Science*, *300*(5624), pp.1393-1393.

Ooi SL, Dann C, 3rd, Nam K, Leahy DJ, Damha MJ, Boeke JD. **2001**. RNA lariat debranching enzyme. Methods Enzym 342: 233-248.

Rearick D, Prakash A, McSweeny A, Shepard SS, Fedorova L, Fedorov A**. 2011**. Critical association of ncRNA with introns. Nucleic Acids Res 39: 2357-2366. Richter JD. 2007. CPEB: a life in translation. Trends Biochem Sci 32: 279- 285.

Ruskin B, Green MR. **1985**. An RNA processing activity that debranches RNA lariats. Science 229: 135-140.

Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE. **2011**. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci*. **108**:20497–20502.

Sverdlov, A.V., Rogozin, I.B., Babenko, V.N. and Koonin, E.V., 2003. Evidence of splice signal migration from exon to intron during intron evolution. *Current Biology*, *13*(24), pp.2170-2174.

Sverdlov, A.V., Rogozin, I.B., Babenko, V.N. and Koonin, E.V., 2005. Conservation versus parallel gains in intron evolution. *Nucleic Acids Research*, *33*(6), pp.1741-1748.

Taggart AJ, Lin C-L, Shrestha B, Heintzelman C, Kim S and Fairbrother WG. **2017**. Large-scale analysis of branchpoint usage across species and cell lines. Genome Res. 27, 639–649.

Verhelst, B., Van de Peer, Y. and Rouzé, P., 2013. The complex intron landscape and massive intron invasion in a picoeukaryote provides insights into intron evolution. *Genome biology and evolution*, *5*(12), pp.2393-2401.

Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, Chen LL. **2012**. Long noncoding RNAs with snoRNA ends. Mol Cell 48: 219- 230.

Wu, B., Macielog, A.I. and Hao, W., 2017. Origin and spread of spliceosomal introns: insights from the fungal clade Zymoseptoria. *Genome Biology and Evolution*, *9*(10), pp.2658-2667.

Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. **2013**. Circular intronic long noncoding RNAs. Mol Cell 51: 792-806.

CHAPTER 1

Stable lariat intronic RNAs in the cytoplasm of *X. tropicalis* oocytes

INTRODUCTION

Recently the Gall Lab carried out a high-throughput sequencing study of RNA from the giant oocyte nucleus or germinal vesicle (GV) of the frog *Xenopus tropicalis* (Gardner *et al*. 2012). They found many intronic sequences in the GV and showed that these intronic sequences are derived from the same strand as the mRNA and probably originate by splicing from pre-mRNA molecules. Because these sequences are stable for at least 48 h, they named them stable intronic sequence RNA or sisRNA.

In this study we show that sisRNAs also occur in the cytoplasm of the *X. tropicalis* oocyte. We are sure of the cytoplasmic localization, because the GV can be removed intact from the oocyte before the cytoplasmic RNA is extracted and sequenced. Cytoplasmic sisRNAs are more abundant on a molar basis than nuclear sisRNAs. We show that cytoplasmic sisRNAs are insensitive to the exonuclease RNase R and that half, if not all, exist as lariat molecules. Cytoplasmic sisRNAs persist after fertilization of the egg until at least the mid-blastula transition, when major transcription first begins in the early embryo. Because of their cellular localization and persistence during early embryogenesis, we suggest that cytoplasmic sisRNAs could play a role in regulating translation of mRNAs.

RESULTS

sisRNA in the cytoplasm

Identification of rare cytoplasmic RNAs depends critically on the ability to isolate a sample of cytoplasm completely free of nuclear contamination. The GV can be removed from an oocyte with a pair of jewelers' forceps in a matter of seconds (Gall and Wu 2010). Because one oocyte contains roughly 1 μg of total RNA, only 5 -10 enucleated oocytes are needed for RNA extraction and library preparation. Furthermore, less than 1% of total cellular RNA resides in the GV (Gardner et al. 2012). Thus, collection of an uncontaminated sample of cytoplasm is a trivial operation that requires only a few minutes. Experimentally, we cannot detect highly abundant nuclear RNAs in our cytoplasmic samples, although they are readily detectable in RNA from whole oocytes (Figs. S1 and S2). Thus, on both theoretical and experimental grounds, we are confident that our cytoplasmic RNA samples are not contaminated with nuclear RNA.

Purified cytoplasmic, nuclear, and whole oocyte RNA samples were depleted for rRNA and subjected to high-throughput sequencing (Fig. 1A). For each sample 60 to 100 million reads (100 bp) were obtained, of which roughly 90% mapped to version 4.1 of the *X. tropicalis* genome (Hellsten et al. 2010). As in earlier experiments (Gardner et al. 2012), cytoplasmic reads mapped primarily to exons, whereas nuclear reads mapped primarily to intronic regions of protein coding genes and to annotated nuclear RNAs (e.g., snRNAs, snoRNAs) (Fig. S1). Roughly 30% of reads in both fractions mapped to unannotated regions of the genome and were not further analyzed.

Of special interest were approximately 3% of cytoplasmic reads that mapped to introns. Two independent samples were sequenced with essentially identical results (Fig. S3). Cytoplasmic intronic reads nearly always mapped as a single peak within a relatively short intron, with no reads crossing the exon-intron boundaries (Fig. 1B). These features suggested that the reads represented independent molecules, not retained introns, which would have reads that cross the exon-intron boundaries.

To confirm this hypothesis, we carried out RT-PCR experiments with exonic and intronic primers to test four introns (Fig. 1C). We first tested the strand specificity of the intronic sequences by using either the forward or reverse primer separately in the RT step. In all four cases only the reverse primer gave a product, showing that the intronic sequences are derived from the same strand as the adjacent exonic sequences. This observation was later confirmed for all cytoplasmic intronic sequences by sequencing the first strand cDNA of the library according to the TruSeq Stranded method from Illumina (Fig. S4). We also tested for sequences that spanned the exon-intron boundary. RT-PCR was carried out with a combination of one exonic and one intronic primer. No products were detected when cytoplasmic RNA was used as the substrate. The same primers were able to amplify from in vitro transcribed RNA that spanned the entire exon-intron-exon region (Fig. 1C). These experiments demonstrate that the cytoplasmic intronic sequences are not attached to the flanking exonic sequences and hence belong to separate molecules. They are, however, transcribed from the same strand as the exonic sequences and could be derived originally from a splicing event.

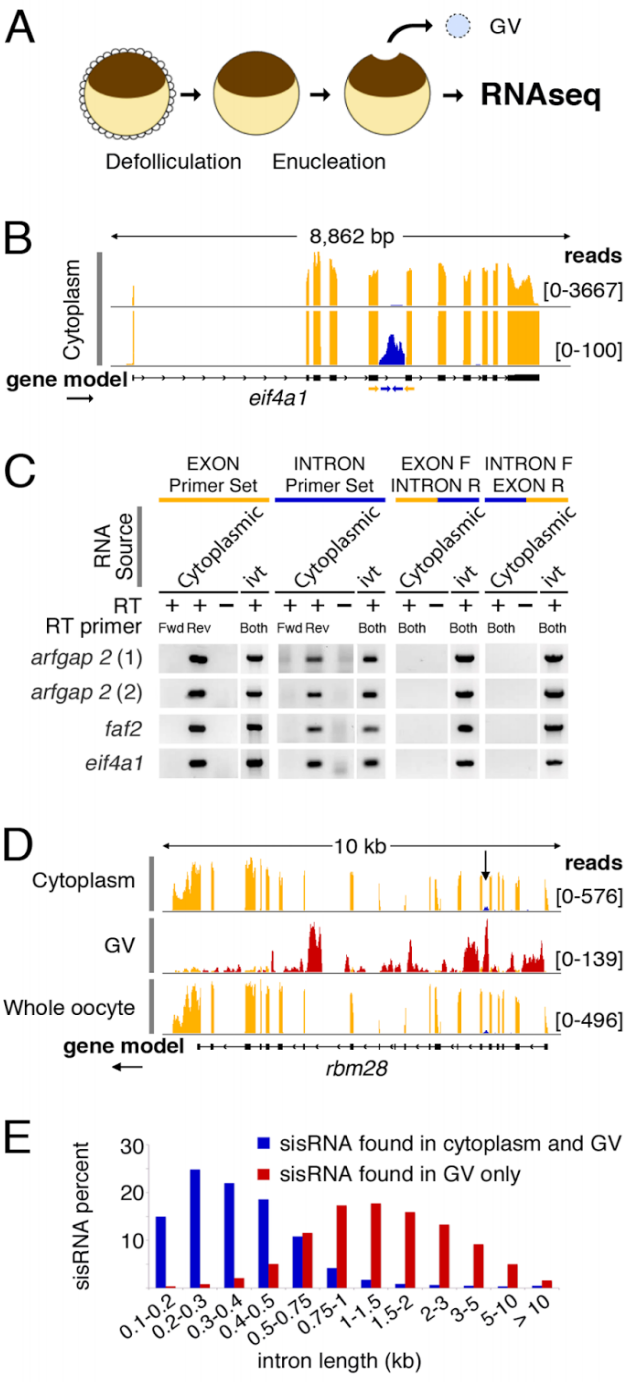


Figure 1. Characterization of cytoplasmic sisRNAs. (A) The oocyte nucleus or germinal vesicle (GV) can be removed from a defolliculated oocyte to provide a sample of cytoplasm uncontaminated with nuclear RNA. (B) IGV browser view of a typical gene bearing a cytoplasmic sisRNA. Intronic reads (blue) occur in one intron of the eif4a1 gene (yellow reads). The lower track shows that the intronic reads do not cross the intron-exon borders. The yellow and blue arrows represent primers used for RT-PCR. (C) RT-PCR results showing that cytoplasmic sisRNAs are derived from the sense strand and do not cross the intron-exon borders. ivt = in vitro transcribed pre-mRNA, used as a control. (D) Comparison of cytoplasmic, nuclear (GV), and whole oocyte RNAs from the rbm28 gene. Cytoplasmic and whole oocyte RNAs are essentially identical (arrow points to cytoplasmic sisRNA, blue reads). Nuclear sisRNAs (red reads) are not detectable in the whole oocyte sample. (E) Distribution of sisRNAs relative to intron length. The majority of cytoplasmic sisRNAs derive from shorter introns, whereas nuclear-specific sisRNAs come from a wide range of longer introns.

To test the stability of the cytoplasmic intronic sequences, we placed oocytes in actinomycin D for 18 hr to inhibit pol II transcription. The efficiency of inhibition was verified by inspecting the loss of transcription loops and pol II staining on the lampbrush chromosomes. As in previous experiments with actinomycin D (Callan 1986), inhibition of transcription was rapid and complete within 1-2 hr. Cytoplasmic RNA was purified from treated and untreated oocytes and high-throughput sequencing was performed after rRNA depletion. We did not see any quantitative differences between treated and untreated oocytes samples, and the ratio between intronic transcripts and the mRNAs with which they are associated did not change (Fig. S5). It is well known that cytoplasmic mRNA in the amphibian oocyte is unusually stable (Davidson 1986). We conclude that the cytoplasmic intronic sequences are also stable, at least during the 18 hr period of actinomycin D treatment.

In summary, intronic sequences in the cytoplasm are derived from the same strand as the mRNA with which they are associated, are independent molecules, and are stable for many hours. In these respects they are similar to sisRNAs from the nucleus and henceforth will be designated cytoplasmic sisRNAs.

Cytoplasmic sisRNAs are more abundant than nuclear sisRNAs and come from shorter introns

Whereas nuclear sisRNAs are derived from roughly half of all introns, cytoplasmic sisRNAs map to only about 5% of introns. Thus cytoplasmic sisRNAs represent a subset of all sisRNAs. Although derived from a limited number of introns,

cytoplasmic sisRNAs are much more abundant on a molar basis than nuclear sisRNAs. This can be deduced by comparing sequence data from nuclear, cytoplasmic, and whole oocyte RNA. Cytoplasmic RNA and whole oocyte RNA are essentially identical with respect to intronic reads. In other words, cytoplasmic sisRNA reads are equally abundant in the two samples but nuclear sisRNA reads are not detectable in the whole oocyte RNA sample (Fig. 1D). On a whole oocyte basis, therefore, cytoplasmic sisRNAs are less abundant than mRNA sequences from the genes in which they are found, but they are more abundant than nuclear sisRNAs derived from the same genes.

These relationships can be assessed semi-quantitatively by comparing the abundance of snRNAs and snoRNAs in the three samples (Fig. S2). snRNA and snoRNA sequences are not detectable in the cytoplasmic fraction, attesting to their strict nuclear localization. They are by far the most abundant sequences in the nuclear sample, and they are readily detectable in RNA derived from the whole oocyte (nucleus plus cytoplasm). Thus, one can compare the abundance of a given snoRNA to nuclear sisRNAs (in the nucleus sample) and to cytoplasmic sisRNAs (in the whole oocyte sample). Such a calculation shows that cytoplasmic sisRNAs are about 10X more abundant than nuclear sisRNAs from the same gene.

Interestingly, introns that have cytoplasmic sisRNAs tend to be short, between 200 and 500 nucleotides in length. In contrast, the vast majority of nuclear sisRNAs are derived from introns that are 500 to 5,000 nucleotides in length (Fig. 1E). We have not detected any other characteristics at the sequence level that differ between nuclear and cytoplasmic sisRNAs.

Cytoplasmic sisRNAs are resistant to RNase R

As a first attempt to characterize cytoplasmic sisRNA molecules in more detail, we carried out 5' and 3' RACE experiments on selected examples. These experiments invariably failed, despite success with appropriate controls performed at the same time. Suspecting that the ends of the cytoplasmic sisRNA molecules might be protected by some modification, we treated cytoplasmic RNA with RNase R. This enzyme is a processive exonuclease that degrades linear single-stranded RNAs from the 3' end regardless of internal secondary structure (Cheng and Deutscher 2002; Cheng and Deutscher 2005). It is not able to degrade circular or lariat RNA and it is most efficient when the terminal ribose is not modified.

We treated cytoplasmic RNA with RNase R or water and tested for the degradation of exonic and intronic sequences by RT-PCR. After RNase R treatment, no exonic product was amplified by RT-PCR for three different genes, whereas sisRNAs from the introns of these genes were still detectable (Fig. 2A). In vitro transcribed sisRNAs were used as controls to show that sequence alone did not prevent the enzyme from working. Because sisRNAs are derived from introns, we strongly suspected that their resistance to RNase R was due to their lariat form; that is, failure to have been debranched after splicing.

To determine whether resistance to RNase R is a characteristic of all cytoplasmic sisRNAs, we treated cytoplasmic RNA samples with RNase R or with water as a control, and carried out high-throughput sequencing. Approximately 70 million reads were obtained for each sample. Based on the RT-PCR experiment, we expected that all mRNA molecules would be sensitive to RNase R and all sisRNAs would be stable. The situation

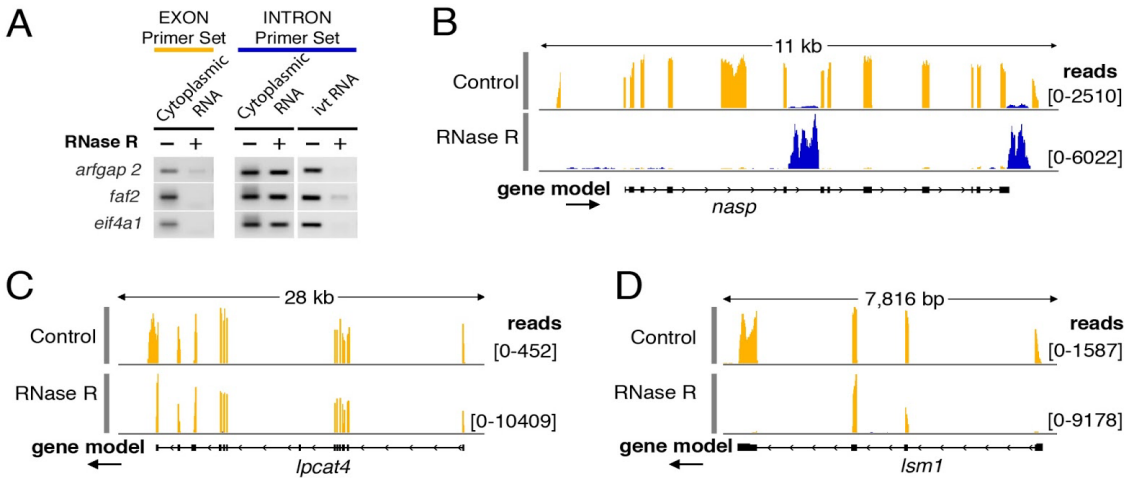


Figure 2. Cytoplasmic sisRNAs are resistant to RNase R. (A) Cytoplasmic RNA was treated with RNase R or water and then used for RT-PCR amplification of exonic and intronic sequences from three genes. Intronic, but not exonic sequences were resistant to RNase R and could still be amplified. The same intronic sequences (ivt RNA), when derived from in vitro transcribed RNA, were digested by RNase R and did not amplify. (B) Cytoplasmic RNA was digested with RNase R and subjected to deep sequencing. The upper track (control) shows that cytoplasmic sisRNA sequences (blue) are rare relative to the coding sequences (yellow) of the nasp gene. The lower track (RNase R treated) shows that the sisRNA sequences are resistant to the enzyme whereas the mRNA is almost completely digested. (C) A relatively rare case in which an mRNA is resistant to RNase R digestion (lpcat4). Presumably digestion began at the 3' end of the mRNA but could not proceed past the middle of the last exon. (D) A circular molecule derived from two exons without the intervening intron in the gene lsm1. Such molecules could arise by "backsplicing," as recently described for human fibroblasts (Jeck et al. 2013).

was somewhat more complex. The total fraction of raw reads that mapped to exons of protein-coding regions dropped from 55% in the control to 20% in the treated sample. Exonic sequences from most mRNAs were either completely undetectable or severely reduced in the sample treated with RNase R, implying that they had been digested efficiently by the enzyme (Fig. 2B). There were interesting exceptions. For example, about 360 mRNAs were enriched in the RNase R sample relative to the control. In most such cases there was a sudden drop in read number near the middle of the last exon at the

3’ end of the molecule (Fig. 2C). A possible explanation for such cases is the presence of a nucleotide with a modified sugar at that position. RNase R, being an exonuclease, could have digested the RNA from the 3’ end until it encountered this nucleotide. In a few cases we saw specific enrichment for 2-4 adjacent exons within a gene (Fig. 2D). Enrichment for numerous exonic sequences after RNase R treatment was recently reported for RNA from cultured human cells and explained as circularization of the molecules by “backsplicing” (Jeck et al. 2013). We did not explore this phenomenon further in our samples.

Conversely, the total fraction of raw reads mapping to introns increased dramatically from 5% to over 40%. Closer analysis of individual genes showed that all sisRNAs seen in the control sample (about 3000) were still represented in the RNase R sample (Fig. 2B and Fig. S6). The patterns of specific sisRNAs on the genome browser did not change after enzyme treatment, suggesting that the molecules were not affected by RNase R. Moreover, because of the greater read-depth of intronic sequences in the treated sample, about 5000 additional sisRNAs were detected that had not been seen in the control.

Cytoplasmic sisRNAs exist as lariats without tails

Definitive evidence that cytoplasmic sisRNAs are lariats without tails comes from several sources. Examination of the average coverage of the most abundant cytoplasmic sisRNAs (top 200) shows a distinct bias toward the 5’ end of the intron (Fig. 3A). Reads map very close to the 5’ splice site but never close to the 3’ splice site, there being a gap

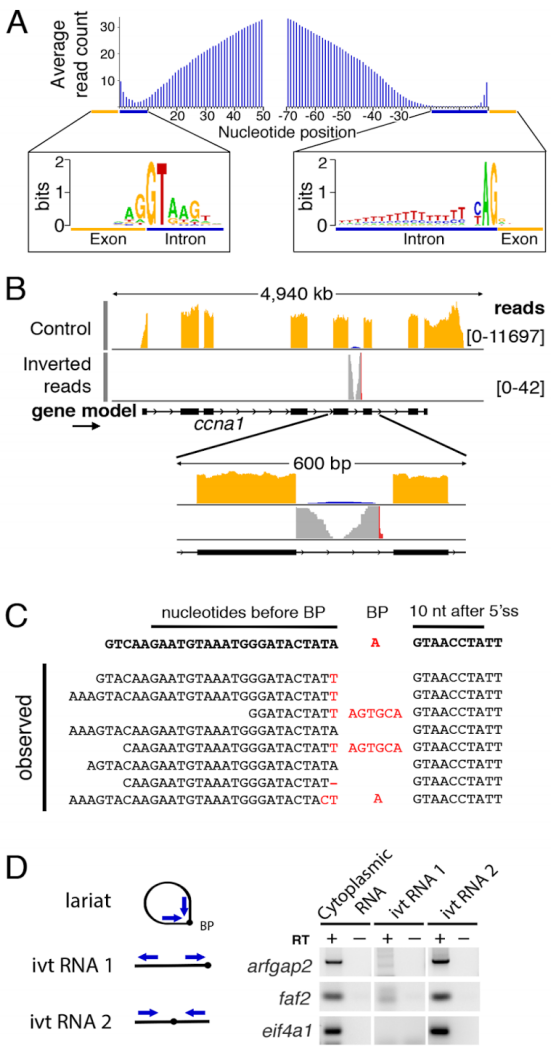


Figure 3. cytoplasmic sisRNAs are circles (lariats without tails). (A) Plot of average read number per nucleotide for 200 sisRNAs relative to nucleotide position within the intron. Reads abut the 5' end of the intron, but are absent from a region of about 30 nucleotides next to the 3' end. Consensus nucleotides are shown for positions near the ends of the introns. (B) Inverted reads contain sequences that cross the splicing branchpoint and can be mapped when appropriately split (Taggart et al. 2012). (C) Some inverted reads contain sequence errors or nucleotides inserted at the branchpoint (red nucleotides), presumably by the reverse transcriptase used for library production. (D) Cytoplasmic sisRNAs yield an RT-PCR product when amplified with outward-facing primers. These same primers amplify in vitro transcribed RNA only when the branchpoint is in the interior of the molecule.

of about 30 nucleotides between the last sisRNA read and the 3' splice site. Because the splicing branch point is typically located at this distance from the 3' splice site, it seemed likely that cytoplasmic sisRNAs are spliced lariats without the 3' tail. Because of their resistance to RNase R, they could be either circles that have not been debranched or linearized lariats with some feature that inhibits RNase R digestion.

To distinguish these alternatives we analyzed unmapped reads from the initial high-throughput sequencing data set. Inverted reads were split and mapped according to the method developed by Taggart et al. (2012). Many inverted reads mapped to intronic regions with a characteristic profile: one end mapped precisely to the 5' splice site, whereas the other end mapped to a region approximately 30 nucleotides upstream of the 3' splice site. Moreover, a variety of "mutations," including short insertions or substitutions, were found at the junction site of the inverted reads. These mutations are shown in red in Fig. 3B and the sequences of a few specific examples are listed in Fig. 3C. It is well known that reverse transcriptase is unreliable when it reaches the 2'-5' linkage at the branch point of a lariat. In most cases it stops at the branch point or it goes through but adds nontemplated nucleotides (Gao et al. 2008). We presume that the errors shown in Fig. 3C were introduced in the reverse transcription step during preparation of the DNA library for high-throughput RNA sequencing. The evidence from inverted reads with mismatches at their junction suggests that cytoplasmic sisRNAs are derived by splicing, but for some reason are not debranched. Instead, they are stabilized in the form of lariats without tails.

High-throughput sequencing data sets contain many PCR artifacts, such as products derived from template switching. Such products can arise from amplification of a heterogeneous template pool with universal primers (Wang and Wang 1996; Acinas et al. 2005). To demonstrate that the inverted reads found in our amplified libraries are not sequencing artifacts but exit in the original sisRNA molecules, we carried out RT-PCR experiments on sisRNAs derived from three genes. First, we used reverse transcriptase to make cDNAs from a cytoplasmic RNA sample using random hexamers. The cDNAs

were amplified by PCR using outwardfacing primers (Fig. 3D, cytoplasmic RNA) and were then cloned and sequenced. The sequences mapped to the region between the 5' end of the intron and the 3' splice site, and there were mutations at the presumed junction site. In vitro transcribed linear sisRNA was not amplified by this method (Fig. 3D, ivt RNA 1), consistent with the primers being specific for inverted sequences. Finally, cDNAs were generated from an inverted in vitro transcribed RNA (Fig. 3D, ivt RNA 2). The sequenced products were similar to those produced from cytoplasmic RNA, except that no mutations were seen at the junction site. We conclude that the inverted reads seen in the high-throughput experiments were not PCR artifacts, but were generated by reverse transcription of lariats in the cytoplasmic RNA sample.

Cytoplasmic sisRNAs are protected from the debranching pathway

The existence of stable RNA lariats in the cytoplasm suggests that cytoplasmic sisRNAs avoid the lariat debranching pathway before their export from the nucleus. We first checked that the RNA lariat debranching enzyme Dbr1 is present in the mature Xenopus oocyte. Western blots of total GV and cytoplasmic extracts show that Dbr1 is present in the GV, as expected, but is not detected in the cytoplasm (Fig. S7).

To test whether cytoplasmic sisRNAs are degradable by debranching activity in the nucleus, we added total cytoplasmic RNA from *X. tropicalis* to cytoplasmic and nuclear extracts from *X. laevis* oocytes (Fig. 4A). The extracts were made from a different species to avoid the complication of endogenous sequences in the extracts. After cytoplasmic RNA from *X. tropicalis* had been incubated in the *X. laevis* nuclear extract

for 2 hr at 25°, *X. tropicalis* mRNAs were still detectable but sisRNAs were not. In contrast, both sisRNAs and mRNAs could be demonstrated after a similar incubation in the cytoplasmic extract (Fig. 4B). These results show that purified cytoplasmic sisRNAs are not inherently resistant to debranching activity. Presumably they are protected in vivo from Dbr1 before they are exported from the nucleus.

An intron giving rise to an abundant cytoplasmic sisRNA can be spliced in the nucleus

Although it is well established that both major and minor splicing occur in the Xenopus oocyte nucleus (Moon et al. 2006; Mereau et al. 2007; Friend et al. 2008), it is possible that cytoplasmic sisRNAs arise by an alternative cytoplasmic pathway. To test this possibility we examined a particularly abundant cytoplasmic sisRNA derived from intron 2 of the faf2 gene of *X. tropicalis*. We made a splicing construct consisting of this intron along with its flanking exons. The capped and polyadenylated RNA construct was injected into an *X. laevis* GV or into an enucleated oocyte, both under mineral oil (Fig. 4C). Two hours later RT-PCR demonstrated that the construct had been spliced in the GV but not in the cytoplasm (Fig. 4D). This experiment demonstrates that an intron that normally gives rise to an abundant cytoplasmic sisRNA can be processed by the nuclear splicing machinery.

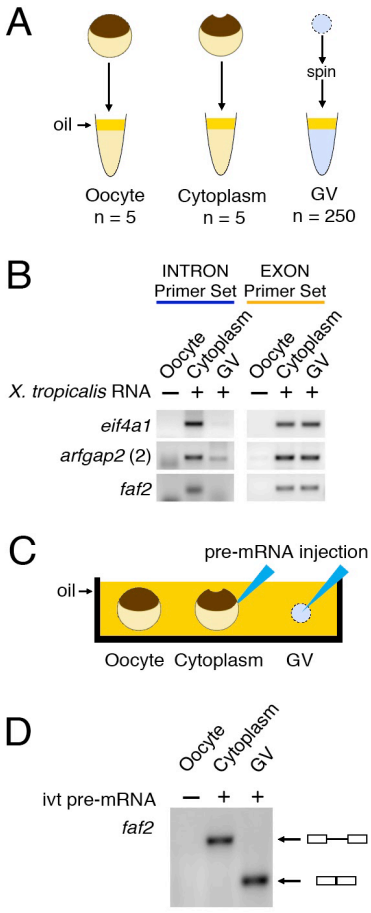


Figure 4. The stability of *X. tropicalis* cytoplasmic sisRNA was tested by RT-PCR after incubation in *X. laevis* cytoplasmic or nuclear extracts. (A) Extracts of *X. laevis* whole oocytes, oocyte cytoplasm, and GVs were made under oil. (B) Cytoplasmic sisRNAs from *X. tropicalis* were degraded by a GV extract from *X. laevis* but not by a cytoplasmic extract (intron primer set, blue). Exonic sequences from *X. tropicalis* were stable for 2 hr in both GV and cytoplasmic extracts (exon primer set, yellow). (C and D) Splicing activity was tested by injecting an in vitro transcribed *X. tropicalis* pre-mRNA construct into the cytoplasm or GV of an *X. laevis* oocyte. (C) Injection of the construct was carried out under mineral oil. (D) After incubation for 2 hr, an RT-PCR reaction was carried out on the cytoplasm and GV using primers from the ends of the construct. The products were run on an agarose gel and stained. Splicing occurred only in the GV, as shown by the expected smaller size of the RT-PCR product.

Some nuclear sisRNAs are also stored as lariats

In our earlier study of GV-specific RNA, we demonstrated the presence of sisRNAs derived from multiple introns of about 6700 protein-coding genes. We did not, however, assess the molecular form of these nuclear sisRNAs. To do so, we analyzed a new sample of GV RNA for branch-point reads (Taggart et al. 2012), as we did for cytoplasmic sisRNAs. We found inverted reads with mismatches at their junction that mapped to the 5′ splice site and upstream of the 3′ splice site in ~20% of the nuclear sisRNAs (Fig. S8A). We interpret these reads as evidence that at least some nuclear sisRNAs exist as lariats despite the presence of active Dbr1 in the GV.

To determine the stability of these lariats we reanalyzed data from an earlier actinomycin D experiment (Gardner et al. 2012). In that experiment we inhibited transcription for 15 h by incubating oocytes in actinomycin D at 20 μg/mL. We then isolated several hundred GVs, extracted the RNA, and performed high-throughput sequencing. Although that sample was not depleted for rRNA, the read-depth was sufficient to detect some inverted intronic reads. A control sample of GV RNA prepared at the same time from untreated oocytes was analyzed with similar results (Fig. S8B). The unusual stability of these lariats implies that at least some nuclear sisRNAs are protected from Dbr1 activity and are not simply transient splicing intermediates. We have evidence from more recent RNase R experiments that the nuclear sisRNA population consists of both lariats and debranched linear molecules.

Accumulation of cytoplasmic sisRNAs during oogenesis

To assess the accumulation of cytoplasmic sisRNAs during oogenesis, we carried out high-throughput sequencing of RNA from three samples of younger oocytes selected on the basis of diameter: under 250 μm, 300-350 μm, and 350-400 μm. Because it would be technically challenging to prepare separate cytoplasm and GV fractions from these oocytes, we compared total RNA with total RNA from mature oocytes (800 μm diameter). Only the most abundant cytoplasmic sisRNAs are detectable in such samples, the majority of reads being derived from mRNA molecules. An example is shown in Fig. 5A. The top track shows cytoplasmic RNA from mature oocytes, demonstrating that the sisRNA under consideration is, in fact, derived from the cytoplasm. RNA from whole oocytes (bottom row, 800 μm) appears essentially identical, showing that there is little or

no contribution to the pattern from the GV. Total RNA from small oocytes (under 250 µm, 300-350 µm, and 350-400 µm) shows very few intronic reads, with more reads in the larger oocytes than in the smaller. A similar pattern of increasing number of reads with oocyte size was apparent when we examined other cytoplasmic sisRNAs individually (Fig. S9).

It was not possible to gather global information on all cytoplasmic sisRNAs in these total oocyte samples. Most reads mapping to annotated introns (about 2% of the total mapped reads) were, in fact, derived from alternative splicing events. On the other hand, certain highly abundant nuclear RNAs, including snoRNAs derived from introns, were detectable in total RNA from oocytes of all sizes. Unlike cytoplasmic sisRNAs, these sequences do not change much in abundance relative to mRNAs derived from the same gene (Fig. 5B).

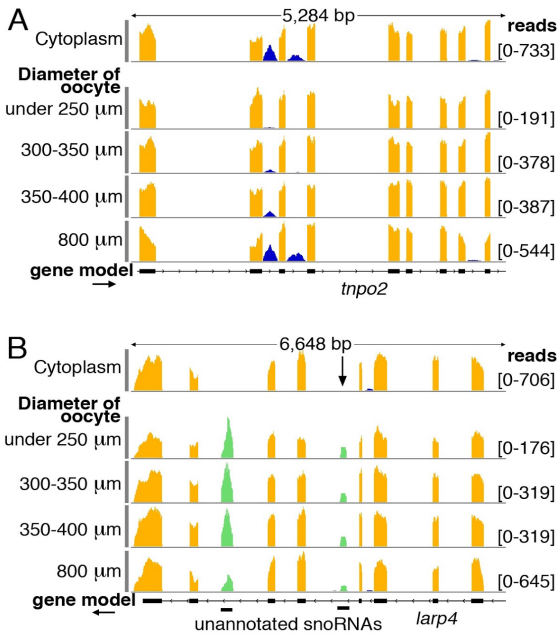


Figure 5. During oogenesis cytoplasmic sisRNAs but not snoRNAs increase in abundance relative to mRNAs. (A) Cytoplasmic sisRNAs (blue) occur in three introns of the tnpo2 gene (top track, cytoplasm from mature oocyte). During oogenesis these sisRNAs increase in abundance relative to exonic sequences (four lower tracks, total RNA from oocytes of increasing size). (B) During oogenesis two nuclear snoRNAs (green) maintain a roughly constant ratio relative to cytoplasmic exonic sequences (yellow). The arrow points to a snoRNA whose size and abundance is similar to that of the sisRNAs detected in (A). The exclusively nuclear location of the snoRNA sequences is shown by their absence from cytoplasm of mature oocytes (top track).

Cytoplasmic sisRNAs persist during early embryogenesis

In our earlier study of sisRNAs we showed by RT-PCR that three abundant sisRNAs persist in the fertilized egg and embryo until at least the blastula stage (Gardner et al. 2012). Unknown to us at that time, all three happen to be cytoplasmic sisRNAs. To gain a more complete picture of sisRNAs during early embryogenesis, we prepared total cell RNA from three stages: eggs after GV breakdown but before fertilization (white spot stage), 4-cell embryos, and early blastulae. Before high-throughput sequencing, half of each sample was digested with RNase R and half was treated with water as a control.

In the controls from the three stages we could detect all cytoplasmic sisRNAs that were evident in a sample of pure cytoplasmic RNA from mature oocytes (in Fig. 6A, compare blue reads in "cytoplasm" with blue reads in "egg," "4-cell,' and "early blastula"). Moreover, the mRNA to sisRNA ratio does not change across these developmental time points. That is, GV breakdown, fertilization, and early development of the embryo have little or no effect on the stability of cytoplasmic sisRNAs. After RNase R treatment of these samples, most mRNA sequences are degraded. Thus, cytoplasmic sisRNAs are dramatically increased relative to exonic sequences (blue intronic vs orange exonic sequences in Fig. 6B for all stages). One now sees nuclear sisRNAs as well (red sequences) in the "egg," "4-cell," and "early blastula" stages. These results suggest that all lariat intronic sequences, both cytoplasmic and nuclear, are stable after GV breakdown until at least the blastula stage. Whether these sisRNAs are nuclear, cytoplasmic, or both during early embryogenesis is not known, since the data necessarily come from unfractionated samples.

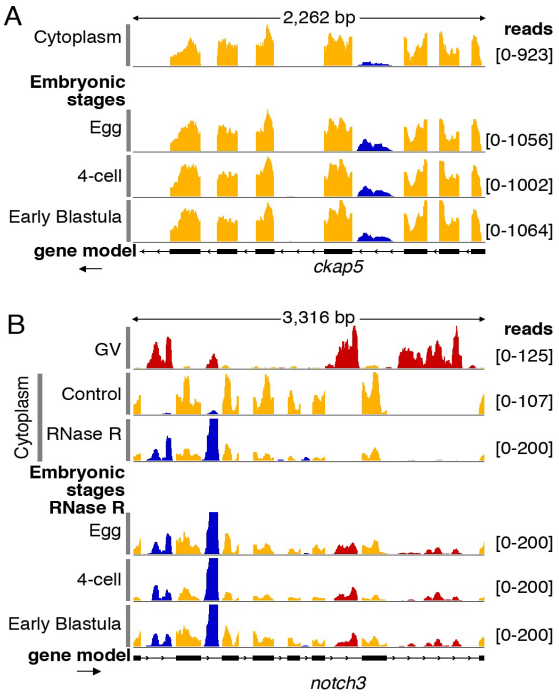


Figure 6. Persistence of sisRNAs in the early embryo. (A) A prominent cytoplasmic sisRNA in the ckap5 gene (top track, blue) persists in the egg, 4-cell, and early blastula stages. (B) The top two tracks show nuclear sisRNAs (red) in the GV and cytoplasmic sisRNAs (blue) in the cytoplasm of mature oocytes. The third track shows that cytoplasmic sisRNAs (blue) are resistant to RNase R relative to the mRNA (yellow). Note that nuclear sisRNAs are not seen in this cytoplasmic sample, attesting to its purity. The bottom three tracks show egg, 4-cell, and early blastula RNA treated with RNase R. These samples show persistence not only of cytoplasmic sisRNAs (blue) but also of nuclear sisRNAs (red) derived from GV breakdown at the time of fertilization.

DISCUSSION

A population of lariat sisRNAs in the cytoplasm

The experiments reported here establish the existence of a population of intronic sequences in the oocyte cytoplasm of *X. tropicalis* and *X. laevis*. Because of their similarity to the stable intronic sequence (sis) RNAs that we previously reported from the oocyte nucleus (Gardner et al. 2012), we refer to these sequences as cytoplasmic sisRNAs. By treating samples with RNase R before sequencing, we identified about 9,000 cytoplasmic sisRNAs derived from 4500 different genes. They come from specific short introns (Fig. 1E), typically only one or two per gene. Further analysis showed that many, if not all, of these molecules are lariats, which probably arise from canonical splicing events (Fig. 3). Cytoplasmic sisRNAs are stable for at least 18 hr after actinomycin treatment. Furthermore, they are transmitted to the embryo at the time of GV breakdown and fertilization, and persist intact until at least the mid-blastula stage. Thus, cytoplasmic sisRNAs constitute a large class of stable lariats derived from a subset of short introns.

By comparing the abundance of molecules in nuclear, cytoplasmic, and whole oocyte RNA samples, it is possible to make rough quantitative estimates of the relative molar concentrations of mRNAs, cytoplasmic sisRNAs, and nuclear sisRNAs (Fig. S2). On average mRNAs are roughly 10X more abundant than cytoplasmic sisRNAs, which in turn are 10X more abundant than nuclear sisRNAs derived from the same gene.

Cytoplasmic lariats derived from introns have been described earlier (Clement et al. 1999; Clement et al. 2001). In the first example, an intron from the mouse T-cell

receptor-beta gene was shown to exist as a lariat, primarily in the nucleus but partly in the cytoplasm. In the second case, three introns from the Pem homeobox gene partitioned preferentially to the cytoplasmic fraction after subcellular fractionation. One intron was further characterized and shown to be a lariat. All three were relatively stable, with half-lives ranging from 9-29 min. An extensively studied example of a cytoplasmic intron is the latency-associated transcript (LAT) derived from an intron in the herpes simplex virus (reviewed in Kent et al. 2003). The cytoplasmic LAT exists in a non-linear form, presumably a lariat (Wu et al. 1996; Rodahl and Haarr 1997). A similar stable intron is derived from the murine cytomegalovirus (Kulesza and Shenk 2006). Finally, in a recent study it was shown that lariat introns accumulate in the cytoplasm of dbr1Δ yeast cells (Armakola et al. 2012). Thus, there is precedence that lariats derived from introns can be found in the cytoplasm, where they are relatively stable.

Stability of RNA classes in the oocyte

Because the ratio of nuclear to cytoplasmic volume drops dramatically during oogenesis (Hausen and Riebesell 1991), quantitative estimates of RNA classes in oocytes of different sizes could be influenced by the relative amounts of nuclear and cytoplasmic RNA. Although nascent transcripts are too low in abundance to be detected even in the smallest oocytes (with the highest nuclear to cytoplasmic volume ratio), certain highly abundant nuclear RNAs, including snoRNAs, are detectable in total RNA from oocytes of all sizes. Figure 5B shows that the ratio of snoRNA reads (from the nucleus) to mRNA reads (from the cytoplasm) remains more or less constant during oocyte development. An mRNA and its associated snoRNA are presumably transcribed in the same event, and

therefore are initially equimolar in amount. The simplest way in which this ratio might remain constant is if their relative stabilities remain constant during oogenesis. Earlier estimates of mRNA stability in *X. laevis* oocytes suggested half-lives of more than 35 days (Davidson 1986). These considerations suggest that snoRNAs are similarly stable.

On the other hand, nuclear sisRNAs in the mature oocyte are present at roughly 1/100 the molar concentration of exonic sequences from the cognate gene (Gardner et al. 2012). Cytoplasmic sisRNAs are intermediate in abundance, being more abundant in the mature oocyte than nuclear sisRNAs from the same gene, but less abundant than exonic sequences. We do not know how these ratios are achieved, because they represent an equilibrium between synthesis and degradation, neither of which has been measured. In principle, both nuclear and cytoplasmic sisRNAs populations could arise because the intron degradation machinery in the GV is inefficient. Alternatively, sisRNAs could represent a subset of sequences that specifically escape the degradation machinery, either by sequestration or other means of protection. In the case of cytoplasmic sisRNAs, physical separation from the nuclear debranching enzyme is a likely component of their higher stability.

Origin of cytoplasmic sisRNAs

At present we have little or no direct information about the origin of cytoplasmic sisRNA molecules. Theoretically they could arise by cytoplasmic splicing, as has been reported for blood platelets (Denis et al. 2005; Schwertz et al. 2006; Rondina et al. 2011) and for neuronal dendrites (Khaladkar et al. 2013; Buckley et al. 2014). Lariats generated by such a mechanism might be partially processed by an exonuclease, but the resulting

lariat molecules would be stable, because the cytoplasm lacks the debranching enzyme Dbr1. We consider this explanation unlikely for several reasons. First, cytoplasmic splicing would involve the excision of retained introns, whereas we do not see extensive evidence in our cytoplasmic samples for unspliced mRNAs or for retention of specific introns that might give rise to cytoplasmic sisRNAs. Second, there is well established evidence for nuclear localization of both major and minor splicing in the Xenopus oocyte (Moon et al. 2006; Mereau et al. 2007; Friend et al. 2008). Third, cytoplasmic sisRNAs use major splice junctions, suggesting that they are processed by conventional nuclear spliceosomes (Fig. 3A). Finally, we confirmed that a pre-mRNA construct for an abundant cytoplasmic sisRNA could be spliced in the nucleus but not in the cytoplasm (Fig. 4, C and D).

We think it more likely that cytoplasmic sisRNAs arise in the GV from incompletely processed introns, even though such a mechanism presents its own theoretical problems. For instance, these spliced introns must avoid the debranching enzyme Dbr1, which we have shown to be present and active in the GV. The fact that many nuclear-specific sisRNAs also exist as lariats is further evidence that nuclear Dbr1 activity is somehow limited to specific intronic transcripts. These sisRNAs could be physically sequestered, so that they do not come in contact with the debranching activity, or they could be in a macromolecular complex that protects them. In situ hybridization can be used to localize introns in the GV and in the cytoplasm, and could give evidence for sequestration within specific nuclear or cytoplasmic bodies. Still another issue for the model proposed here is the mechanism of export from the nucleus. Although the export of various classes of RNA from the nucleus has been extensively investigated (Kohler

and Hurt 2007), the mechanism by which excised introns are exported has not, to our knowledge, been described. Further insight into these issues may be gained by studying the proteins with which lariat introns are associated (both nuclear and cytoplasmic).

Biological role of cytoplasmic sisRNAs

Cytoplasmic and nuclear sisRNAs join an ever-growing list of non-coding RNAs whose specific functions remain to be discovered (Ulitsky and Bartel 2013; van Heesch et al. 2014). A recent study identified numerous circular intronic sequences in cultured human cell lines (HeLa and H9) (Zhang et al. 2013). The authors show by in situ hybridization that at least one of these circular RNAs, derived from an intron in ankrd52, associates with its site of transcription in the nucleus. Furthermore, it interacts with the pol II machinery and acts as a positive regulator of transcription. Many or most of these circular RNAs in human cells reside in the nucleus and thus may correspond to the sisRNAs we found in the Xenopus oocyte nucleus (Gardner et al. 2012). However, because of their cytoplasmic localization, the sisRNAs described here cannot have a direct role at their site of transcription. Without further evidence, only the most general arguments can be made about their function(s). Two features stand out. First, they are derived from a relatively small number of specific introns, implying that they are not random sequences derived from an inefficient degradation machinery. Second and more importantly, they are found in the cytoplasm long before nuclear envelope breakdown at the time of oocyte maturation. Given that these specific introns are produced by the nuclear splicing machinery, there must be a mechanism for exporting them out of the nucleus. The existence of such a mechanism implies that cytoplasmic localization is key

to their function. It is well known that protein synthesis in the oocyte and fertilized egg is highly regulated (Richter 2007; Richter and Lasko 2011) and cytoplasmic sisRNAs could be one of the players in this regulation. For instance, cytoplasmic sisRNAs might be associated with their cognate mRNAs and in some way help regulate their translation. Single molecule in situ hybridization, using probes for the two RNAs, could provide definitive evidence for or against this hypothesis. More generally, cell fractionation should be helpful in suggesting a function for cytoplasmic sisRNAs. Association of cytoplasmic sisRNAs with either polysomes or single ribosomes would be highly suggestive of a regulatory role (positive or negative) in protein synthesis.

MATERIALS AND METHODS

Animals, embryos, and oocytes

Female *X. tropicalis* of various ages were purchased from Xenopus 1. Animals were anesthetized with 0.15% tricaine methane sulfonate, and one or both ovaries were removed surgically. Pieces of ovary were cultured in OR2 medium (Wallace et al. 1973) at room temperature for up to several days. Follicle cells were removed from oocytes with collagenase (Liberase, Roche Applied Science). Success of the treatment was verified by staining with DAPI (4', 6-diamidino-2-phenylindole) at 1 μg/mL and examining under low magnification in a fluorescence microscope (Simeoni et al. 2012). Transcription was inhibited by incubating pieces of ovary in actinomycin D (20 μg/mL) in OR2 medium. Embryos were obtained by in vitro fertilization. Females and males were primed with 100 units of human chorionic gonadotropin (HCG) the night before use. The next day, females were injected with 400 units of HCG and males with 200 units. Four hours later, females were squeezed to obtain eggs, whereas testes were removed from the males and macerated in saline. The sperm suspension was dispersed onto the freshly squeezed eggs. Eggs and embryos were manually dejellied and collected with minimal liquid in 1.5-mL Eppendorf tubes on dry ice.

Preparation of nuclear and cytoplasmic RNA fractions

GVs were manually isolated in an isotonic saline solution at pH 5.6–5.8 (83 mM KCl, 17 mM NaCl, 6.0 mM Na2HPO4, 4.0 mM KH2PO4, 1 mM MgCl2, 1.0 mM dithiothreitol, adjusted to pH 5.6–5.8 with HCl) (Gardner et al. 2012). The nuclear envelope was removed with jewelers' forceps and the GV contents were collected in 10

mM sodium citrate and 5 mM EDTA (pH 5.0) for storage. Depending on the experiment, up to 500 GVs were collected over a period of several days. Samples of cytoplasm were obtained by removing the GV from defolliculated oocytes (Simeoni et al. 2012) and immediately placing the enucleated oocyte into a 1.5-mL Eppendorf tube on dry ice. Adequate cytoplasmic RNA could be obtained from 5-10 enucleated oocytes. RNA was extracted with TRIzol reagent (Ambion) and purified with the Direct-zol kit (Zymo Research). RNA was quantitated with a NanoDrop 2000 spectrophotometer (Thermo Scientific) and further characterized with a Bioanalyzer 2100 (Agilent).

Library preparation, sequencing, and sequence analysis

Samples were depleted of rRNA using Ribozero or Ribozero-Gold kits (Epicenter) according to the standard protocol. Subsequently, libraries were prepared using the TruSeq RNA sample preparation protocol (Illumina) or TrueSeq stranded total RNA sample preparation (Illumina). Sequencing was performed on an Illumina HiSeq 2000 sequencer with either 50 or 100 bp single-end reads. Reads were aligned to the *X. tropicalis* genome (version 4.1) or the *X. laevis* genome (version 6.0) using the TopHat (version 2.0.7) and Bowtie (version 2.0.6.0) sequence alignment programs (Langmead et al. 2009; Trapnell et al. 2009). Exonic and Intronic regions were quantified using BEDtools version 2.15.0 and Cufflinks version 2.0.2 (Quinlan and Hall 2010; Trapnell et al. 2010). Features with FPKM ≥ 2 were considered expressed. Sequence alignments were examined in the Integrative Genomics Viewer (IGV) from the Broad Institute (Robinson et al. 2011). To search for lariats or other circular RNA reads, we ran the

PERL script findlariat.pl (Taggart et al. 2012). Further analysis of sisRNAs was done using BEDtools in conjunction with custom PERL scripts.

RT–PCR analysis

We performed RT-PCR against a few candidates using the one-step RT-PCR kit (Qiagen) to validate bioinformatic predictions. In vitro transcribed RNAs were used as positive control to demonstrate the competence of the primers. To generate DNA templates for in vitro transcription, the regions of interest were amplified from genomic DNA by PCR. A T3 or T7 promoter was included in forward primers. The RNA was transcribed using T3 or T7 RNA polymerase, treated with DNase, and purified with G50 columns according to standard protocols. To confirm the orientation of selected introns, we used a single primer (either forward or reverse primer) for the RT step, followed by PCR with both sets of primers (35 cycles). For detection of lariats, total RNA was reverse-transcribed for 1 h using Episcript (Epicenter) and random hexamers. Either Episcript or SuperScript II (Life Technologies) was essential for reverse transcription through the branch point. Subsequently, PCR was performed using outward facing primer pairs (40 cycles). PCR fragments were cloned with pGEM-T Easy vector system (Promega) and sequenced to verify the specificity of these reactions. Primers used in this study are listed in Table S1 (Appendix).

RNase R

rRNA-depleted samples were denatured at 65°C for 5 min and cooled on ice for 1 min. 10X RNase R buffer, RNasin (1 μL) and either 20 units of RNase R (Epicenter) or

water was then added. The reaction was carried out overnight at 37°C. RT-PCR was carried out directly after the incubation, whereas libraries were made after precipitation of the RNA.

Western blots

For western blots of the RNA lariat debranching enzyme, Dbr1, we used an affinity-purified rabbit polyclonal antibody raised against a region within amino acids 68-324 of human Dbr1 (Abcam, ab154230). The antibody was used at 1:10,000 in 1% blocking reagent (Roche). The secondary antibody was anti-rabbit IgG conjugated with horseradish peroxidase, used at a dilution of 1:20,000 in 5% nonfat dry milk. Peroxidase was detected with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific).

Debranching assay

Xenopus oocyte nuclei remain physiologically active for hours when isolated under mineral oil (Paine et al. 1992; Yu et al. 2001; Deryusheva and Gall 2004). We prepared both nuclear and cytoplasmic extracts from *X. laevis* oocytes and tested for debranching activity. Two hundred and fifty nuclei were collected under oil and centrifuged at 20,800 g for 2 min to obtain a clear nuclear extract. Cytoplasm from 5 oocytes was gently pipetted to obtain a cytoplasmic homogenate. Purified cytoplasmic RNA from *X. tropicalis* oocytes was added to the extracts, incubated for 2 hr at 25°C, and purified according to the RNAqueous micro-kit protocol (Ambion).

Splicing assay

Splicing was assayed in single GVs or single enucleated oocytes of *X. laevis* (under oil). Each was injected with a capped and polyadenlylated RNA construct consisting of intron 2 of the faf2 gene of *X. tropicalis* along with its flanking exons. After incubation for 2 hr at 25°C, RNA was purified using the RNAqueous micro-kit protocol (Ambion). Spliced and unspliced products were detected by RT-PCR using primers from the ends of the exons.

REFERENCES

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. **2005**. PCRinduced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol 71: 8966-8969.

Armakola M, Higgins MJ, Figley MD, Barmada SJ, Scarborough EA, Diaz Z, Fang X, Shorter J, Krogan NJ, Finkbeiner S, Farese RV Jr, and Gitler AD. **2012**. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. Nature Genet 44: 1302-1309.

Buckley PT, Khaladkar M, Kim J, Eberwine J. **2014**. Cytoplasmic intron retention, function, splicing, and the sentinel RNA hypothesis. Wiley Interdiscip Rev RNA 5: 223-230.

Callan HG. **1986**. Lampbrush Chromosomes. Springer-Verlag, Berlin.

Cheng Z-F, Deutscher MP. **2005**. An Important role for RNase R in mRNA decay. Mol Cell 17: 313-318.

Cheng ZF, Deutscher MP. **2002**. Purification and characterization of the Escherichia coli exoribonuclease RNase R. Comparison with RNase II. J Biol Chem 277: 21624-21629.

Clement JQ, Maiti S, Wilkinson MF. **2001**. Localization and stability of introns spliced from the Pem homeobox gene. J Biol Chem 276: 16919-16930.

Clement JQ, Qian L, Kaplinsky N, Wilkinson MF. **1999**. The stability and fate of a spliced intron from vertebrate cells. RNA 5: 206-220.

Davidson EH. **1986**. Gene Activity in Early Development. Academic Press, Orlando, FL.

Denis MM, Tolley ND, Bunting M, Schwertz H, Jiang H, Lindemann S, Yost CC, Rubner FJ, Albertine KH, Swoboda KJ et al. **2005**. Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. Cell 122: 379-391.

Deryusheva S, Gall JG. **2004**. Dynamics of coilin in Cajal bodies of the Xenopus germinal vesicle. Proc Natl Acad Sci USA 101: 4810-4814.

Friend K, Kolev NG, Shu MD, Steitz JA. **2008**. Minor-class splicing occurs in the nucleus of the Xenopus oocyte. RNA 14: 1459-1462.

Gall JG, Wu Z. **2010**. Examining the contents of isolated Xenopus germinal vesicles. Methods 51: 45-51.

Gao K, Masuda A, Matsuura T, Ohno K. **2008**. Human branch point consensus sequence is yUnAy. Nucleic Acids Res 36: 2257-2267.

Gardner EJ, Nizami ZF, Talbot CC, Jr., Gall JG. **2012**. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of Xenopus tropicalis. Genes Dev 26: 2550-2559.

Hausen P, Riebesell M. **1991**. The early development of Xenopus laevis. An atlas of the histology. Springer-Verlag, Berlin.

Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L et al. **2010**. The genome of the Western clawed frog Xenopus tropicalis. Science 328: 633-636.

Hoskins AA, Moore MJ. **2012**. The spliceosome: a flexible, reversible macromolecular machine. Trends Biochem Sci 37: 179-188.

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. **2013**. Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA 19: 141-157.

Kent JR, Kang W, Miller CG, Fraser NW. **2003**. Herpes simplex virus latency-associated transcript gene function. J Neurovirol 9: 285-290.

Khaladkar M, Buckley PT, Lee MT, Francis C, Eghbal MM, Chuong T, Suresh S, Kuhn B, Eberwine J, Kim J. **2013**. Subcellular RNA sequencing reveals broad presence of cytoplasmic intron-sequence retaining transcripts in mouse and rat neurons. PLoS One 8: e76194.

Kohler A, Hurt E. **2007**. Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol 8: 761-773.

Kulesza CA, Shenk T. **2006**. Murine cytomegalovirus encodes a stable intron that facilitates persistent replication in the mouse. Proc Natl Acad Sci USA 103: 18302-18307.

Langmead B, Trapnell C, Pop M, Salzberg SL. **2009**. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

Mereau A, Le Sommer C, Lerivray H, Lesimple M, Hardy S. **2007**. Xenopus as a model to study alternative splicing in vivo. Biol Cell 99: 55-65.

Moon KH, Zhao X, Yu YT. **2006**. Pre-mRNA splicing in the nuclei of Xenopus oocytes. Methods Mol Biol 322: 149-163.

Ooi SL, Dann C, 3rd, Nam K, Leahy DJ, Damha MJ, Boeke JD. **2001**. RNA lariat debranching enzyme. Methods Enzym 342: 233-248.

Paine PL, Johnson ME, Lau YT, Tluczek LJ, Miller DS. **1992**. The oocyte nucleus isolated in oil retains in vivo structure and functions. Biotechniques 13: 238-246.

Quinlan AR, Hall IM. **2010**. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.

Richter JD, Lasko P. **2011**. Translational control in oocyte development. Cold Spring Harbor Persp Biol 3: a002758.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. **2011**. Integrative genomics viewer. Nature Biotech 29: 24-26.

Rodahl E, Haarr L. **1997**. Analysis of the 2-kilobase latency-associated transcript expressed in PC12 cells productively infected with herpes simplex virus type 1: evidence for a stable, nonlinear structure. J Virol 71: 1703-1707.

Rondina MT, Schwertz H, Harris ES, Kraemer BF, Campbell RA, Mackman N, Grissom CK, Weyrich AS, Zimmerman GA. **2011**. The septic milieu triggers expression of spliced tissue factor mRNA in human platelets. J Thromb Haemost 9: 748-758.

Schwertz H, Tolley ND, Foulks JM, Denis MM, Risenmay BW, Buerke M, Tilley RE, Rondina MT, Harris EM, Kraiss LW et al. **2006**. Signaldependent splicing of tissue factor pre-mRNA modulates the thrombogenicity of human platelets. J Exp Med 203: 2433-2440.

Simeoni I, Gilchrist MJ, Garrett N, Armisen J, Gurdon JB. **2012**. Widespread transcription in an amphibian oocyte relates to its reprogramming activity on transplanted somatic nuclei. Stem Cells Dev 21: 181-190.

Taggart AJ, DeSimone AM, Shih JS, Filloux ME, Fairbrother WG. **2012**. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat Struct Mol Biol 19: 719-721.

Trapnell C, Pachter L, Salzberg SL. **2009**. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. **2010**. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotech 28: 511- 515.

Ulitsky I and Bartel DP. **2013**. lincRNAs: genomics, evolution, and mechanisms. Cell 154: 26-46.

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, Macinnes AW, Cuppen E, Simonis M. **2014**. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol 15: R6.

Wallace RA, Jared DW, Dumont JN, Sega MW. **1973**. Protein incorporation by isolated amphibian oocytes: III. Optimum incubation conditions. J Exp Zool 184: 321-333.

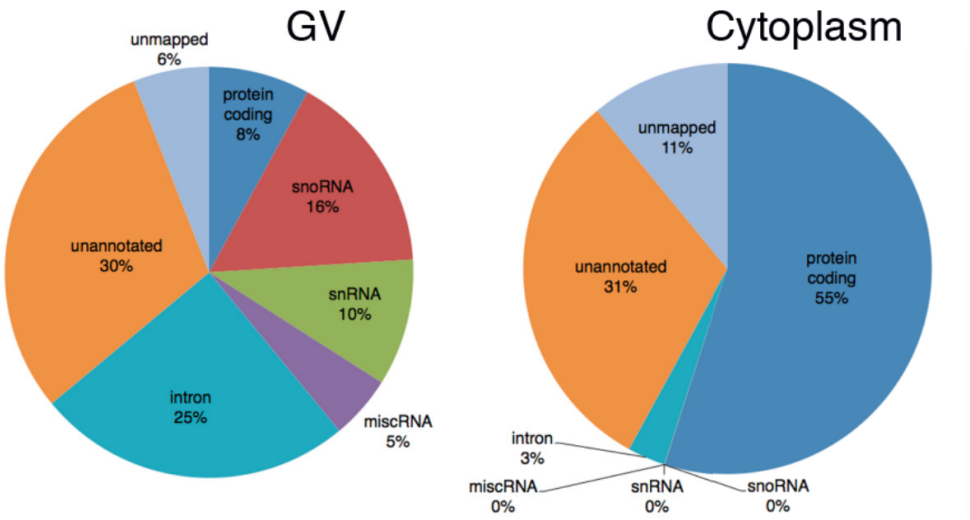
Wang GC, Wang Y. **1996**. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiol 142 (Pt 5): 1107-1114.

Wu TT, Su YH, Block TM, Taylor JM. **1996**. Evidence that two latency associated transcripts of herpes simplex virus type 1 are nonlinear. J Virol 70: 5962-5967.

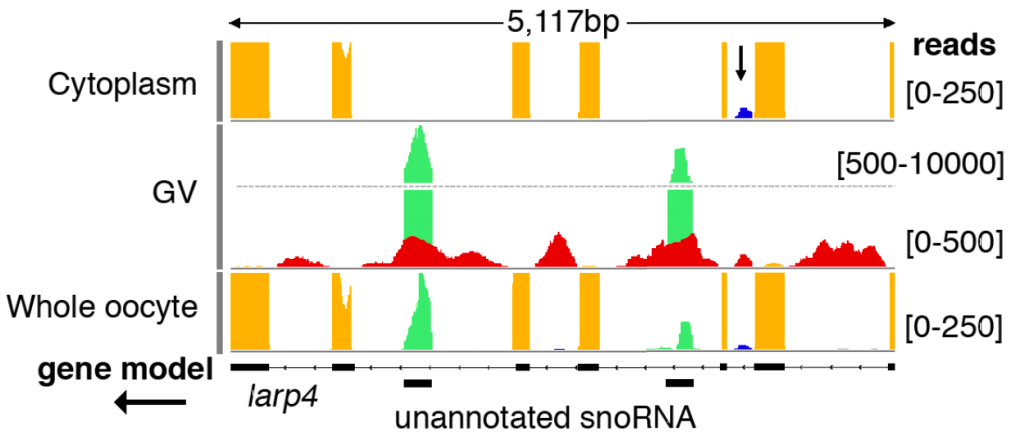
Yu YT, Shu MD, Narayanan A, Terns RM, Terns MP, Steitz JA. **2001**. Internal modification of U2 small nuclear (sn)RNA occurs in nucleoli of Xenopus oocytes. J Cell Biol 152: 1279-1288.

Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. **2013**. Circular intronic long noncoding RNAs. Mol Cell 51: 792-806.

Supplementary Figures



Supplemental Figure S1. Relative numbers of reads from an experiment in which RNA samples from *X. tropicalis* oocyte nuclei (GV) and cytoplasm were subjected to deep sequencing. In the nucleus snRNAs, snoRNAs, and intronic sequences make up the majority of annotated sequences. In the cytoplasm, mRNA sequences constitute the major class of annotated sequences, with intronic sequences next in abundance. Noteworthy is the complete absence of snoRNA and snRNA sequences from the cytoplasm, attesting to the lack of nuclear contamination in this fraction.



Supplemental Figure S2. Use of a snoRNA as an internal standard to compare relative abundance of cytoplasmic and nuclear sisRNAs. snoRNA sequences (green) are strictly nuclear, as shown by their absence from a sample of cytoplasmic RNA (top track). They are sufficiently abundant to appear in the whole oocyte sample (bottom track) and they are by far the most abundant sequences in the GV sample (middle track, note the change in scale to accommodate the extremely abundant snoRNA). Thus snoRNA abundance (green) can be compared to cytoplasmic sisRNA abundance (blue) in the whole oocyte sample and to nuclear sisRNA abundance (red) in the GV. From these ratios the relative abundance of cytoplasmic sisRNA to nuclear sisRNA can be approximated, in this case

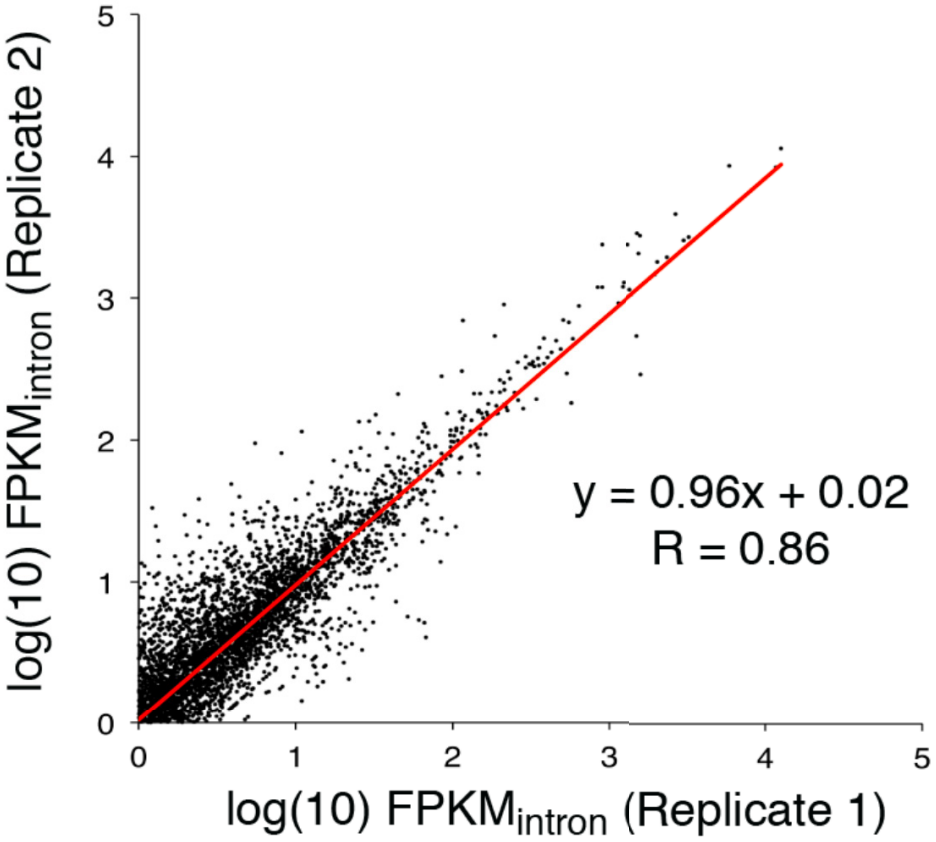
18.9/128.6 X 5604/80.9 = 10.2

18.9 reads/100bp = sisRNA in cytoplasm (blue in whole oocyte)

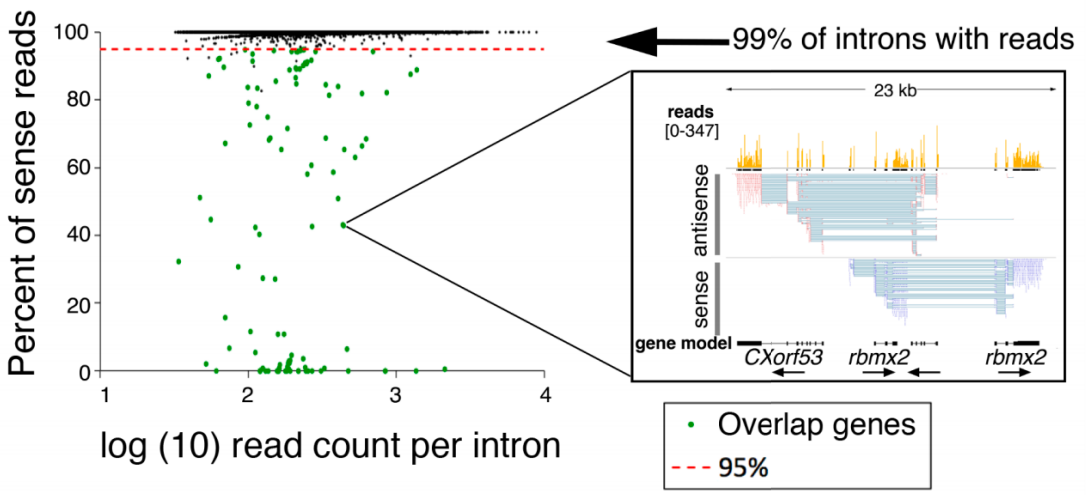
128.6 reads/100bp = average of the 2 snoRNAs (green in whole oocyte)

5603.9 reads/100bp = average of the 2 snoRNAs (green in GV)

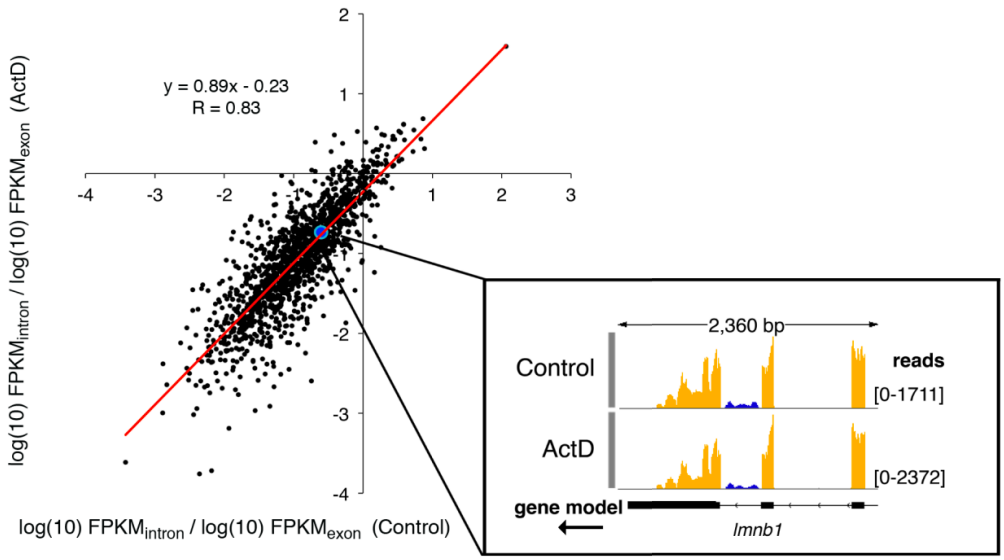
80.9 reads/100bp = average of the nuclear sisRNAs (red in GV)



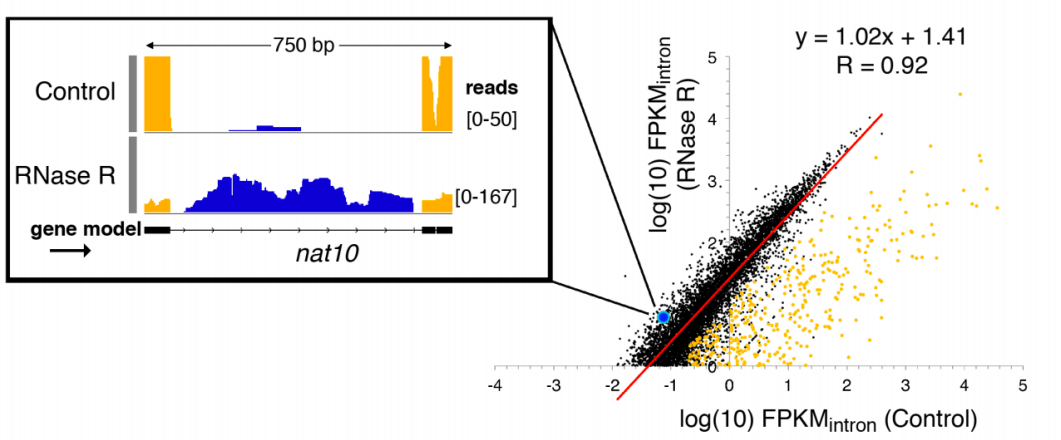
Supplemental Figure S3. Comparison of $\log_{10}$ $FPKM_{intron}$ values for 4417 introns from two independent samples of cytoplasmic RNA. Reproducibility between experiments was high (R = 0.86). These introns were derived from a total of 2963 genes, approximately 30% of all genes with an $FPKM_{exon} \geq 2$.



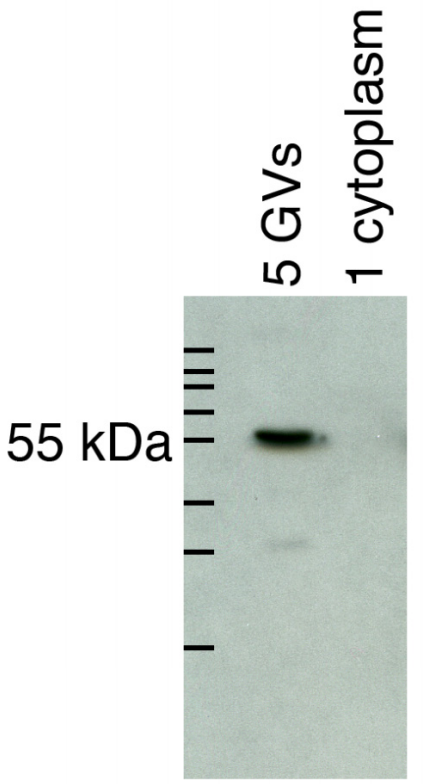
Supplemental Figure S4. All cytoplasmic intronic reads are from the same strand as corresponding exonic reads. Reads were mapped to 14,875 introns in this sample of cytoplasmic RNA. One percent of these (green dots) appeared to have some anti-sense reads. When examined individually, however, each case involved overlapping genes with opposite orientations. The inset shows one such example.



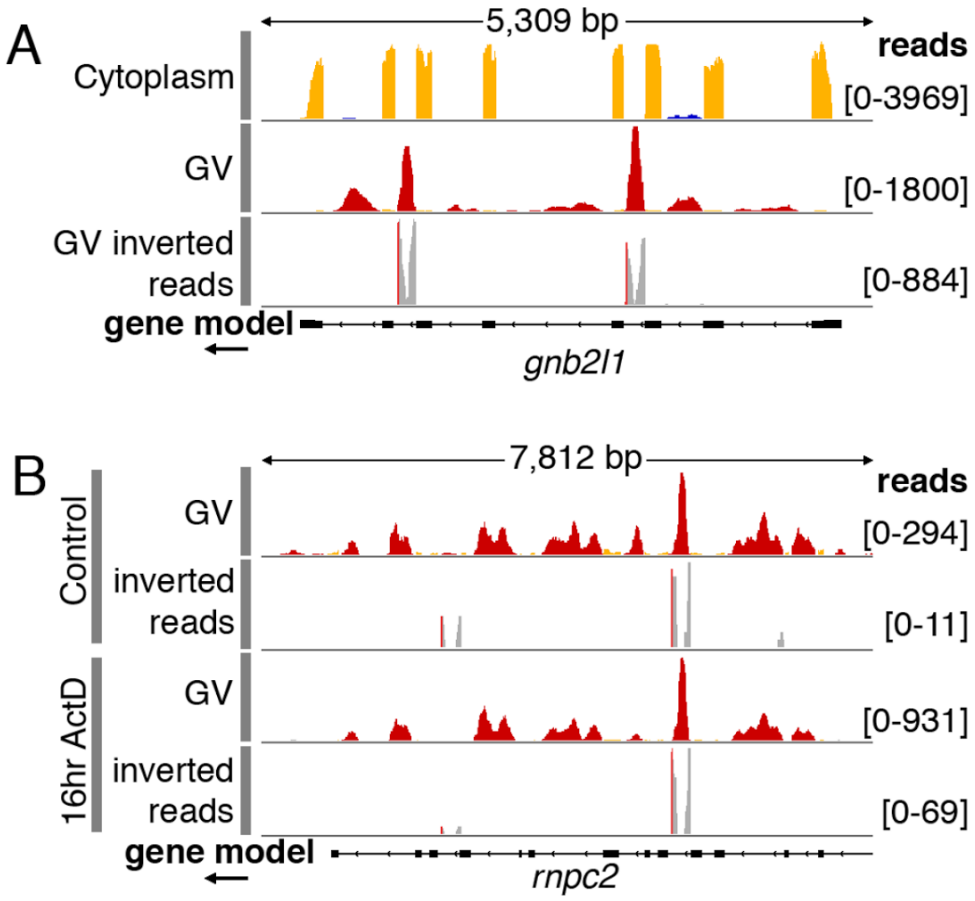
Supplemental Figure S5. Cytoplasmic sisRNA sequences are as stable as mRNA sequences after inhibition of transcription with actinomycin D. Shown is a comparison of $\log_{10}$ $FPKM_{intron}$ / $\log_{10}$ $FPKM_{exon}$ values for cytoplasmic RNA from control and actinomycin-treated oocytes. Correlation between the two samples is high (R = 0.83). Each point represents the ratio between the $\log_{10}$ FPKM value for a specific intron and the $\log_{10}$ FPKM value for the entire mRNA. The inset shows an example: the intron value is based on the blue fragments and the intron length, the exon value on the yellow fragments and the total exonic length.



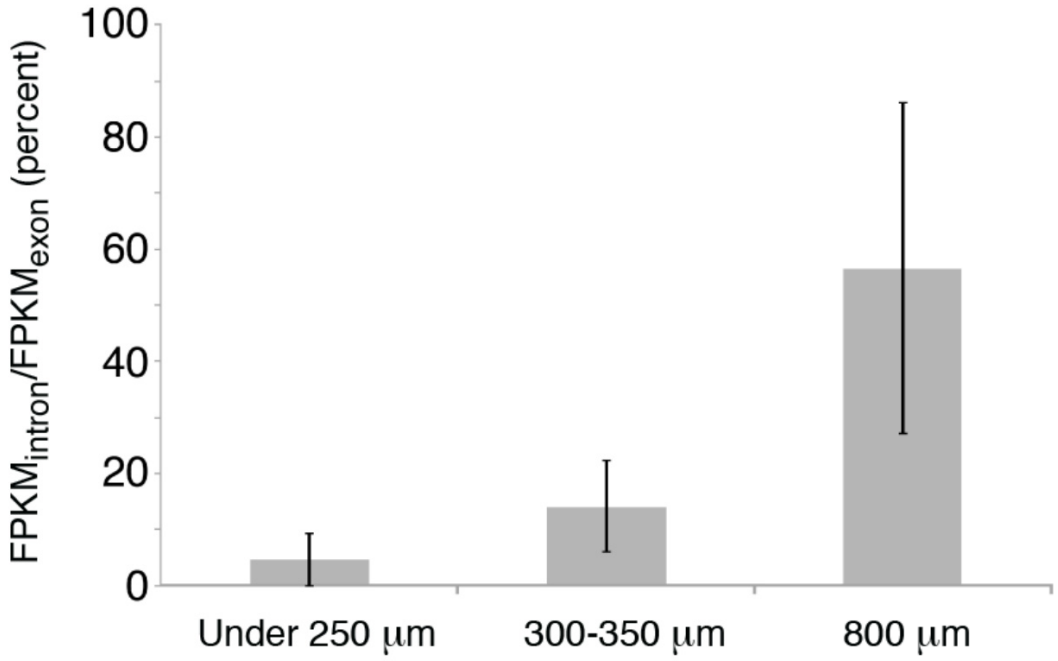
Supplemental Figure S6. Treatment of a cytoplasmic RNA sample with RNase R increases the number of sisRNA reads by 10-100 fold. Log_{10} $FPKM_{intron}$ values for control and RNase R treated samples are plotted. The y-intercept of the regression line is 1.4, equal to an average 25 fold increase in reads after RNase R. Points that fell away from the main cluster (yellow) were examined individually and in every case corresponded to alternative splicing, not genuine sisRNA sequences. Inset shows the increase in reads (blue) for a typical sisRNA (in the *nat10* gene).



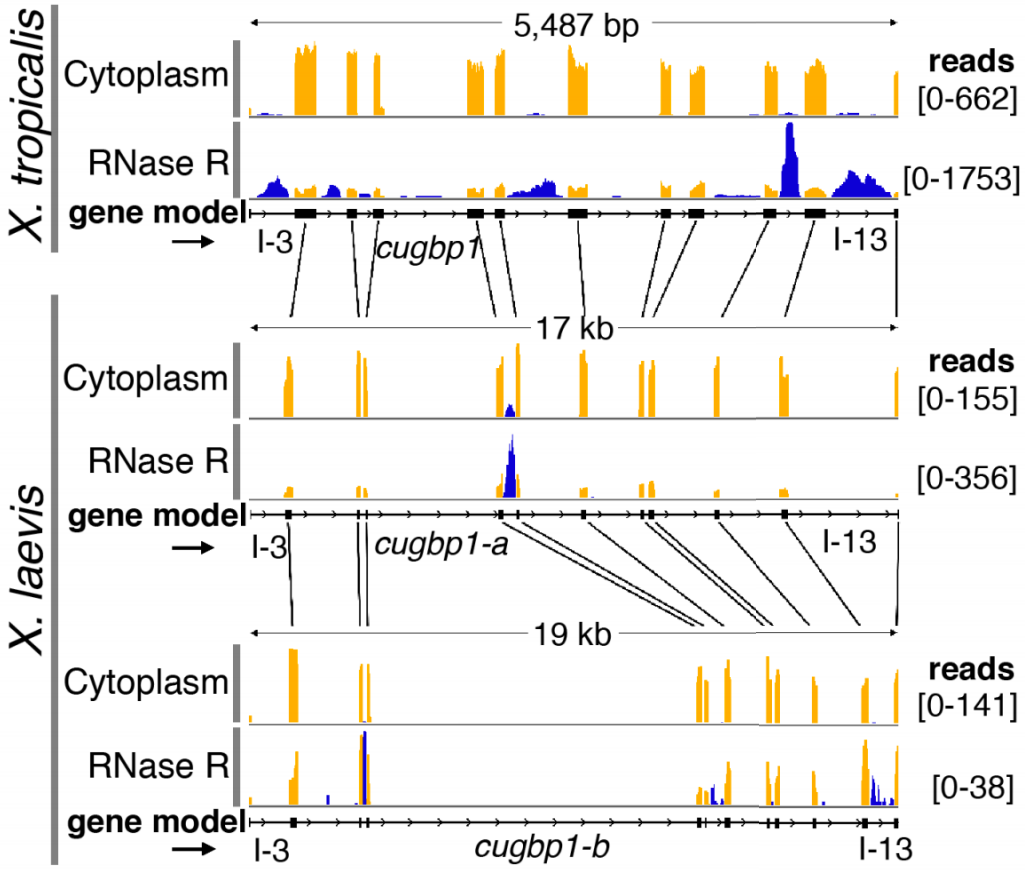
Supplemental Figure S7. The RNA lariat debranching enzyme Dbr1 is present in the GV but is not detectable in the cytoplasm of the *X. tropicalis* oocyte. Shown here is a western blot of proteins from 5 GVs or 1 cytoplasm probed with an antibody against human Dbr1.



Supplemental Figure S8. Some nuclear sisRNAs are stable lariats. *(A)* Inverted reads with mismatches at the junction can be found for some nuclear sisRNA molecules by searching the pool of unmapped reads (Taggart et al. 2012). These sisRNA molecules are presumably lariats. The sisRNAs that lack inverted reads could still be lariats, since inverted reads are produced only if the reverse transcriptase can pass through the branchpoint. *(B)* All nuclear sisRNAs are equally stable after transcription has been inhibited by actinomycin D. The persistence of lariat sisRNAs after actinomycin treatment shows that they are not transient splicing intermediates.



Supplemental Figure S9. The abundance of cytoplasmic sisRNAs increases relative to mRNA during oogenesis. The graph shows the averages for the 15 most abundant cytoplasmic sisRNAs.



Supplemental Figure S10. Cytoplasmic sisRNAs do not occupy the same introns in orthologous genes of *X. tropicalis* and *X. laevis*. The top two tracks show that the *cugbp1* gene (introns 3-13) of *X. tropicalis* gives rise to multiple cytoplasmic sisRNAs (blue), with the most prominent in intron 12. The two homologous genes in *X. laevis* (*cugbp1-a* and *cugbp1-b*) differ from each other and from the *X. tropicalis* gene. The most prominent sisRNA comes from intron 7 in *cugbp1-a* (middle tracks) but from intron 5 in *cugbp1-b* (bottom tracks).

CHAPTER 2

Stable lariat intronic RNAs in the cytoplasm of vertebrate cells

INTRODUCTION

Although intronic RNAs are typically degraded within minutes, high throughput sequencing revealed the existence of unusually abundant intronic RNAs in *Xenopus* (Gardner *et al.*, 2012) and in cultured human cells (Zhang *et al*., 2013). In the frog oocyte, many intronic RNAs are highly stable; hence, their name stable intronic sequence RNA or sisRNA. Notably, the most abundant sisRNAs are stored as circular molecules (lariats without a tail) and, unexpectedly, are exported to the cytoplasm (Chapter 1). Until now, it has remained unknown whether cytoplasmic sisRNAs are unique to frog oocytes.

In this study, I searched for abundant intronic RNAs in other species and found circular sisRNAs in human, mouse, chicken, frog, and zebrafish cultured cells and diverse tissues. I also found sisRNAs in mammalian red blood cells (RBCs). Because mammalian RBCs lack nuclei, this observation was direct evidence for cytoplasmic sisRNAs. I discuss functions that cytoplasmic sisRNAs may perform independently from their host genes.

RESULTS

Stable intronic sequence RNAs are detectable in cell culture

In my previous study (Chapter 1), cytoplasmic sisRNAs were enriched by treating RNA with RNase R, a 3'-5' exonuclease that degrades most linear molecules. In other words, the enrichment of circular RNA is sufficient to identify sisRNA without any additional selection.

To search for sisRNA in other species, I started with mouse cell culture. I treated 3T3 cells with α-amanitin to inhibit transcription and allow degradation of transient intronic RNA (untreated cells served as controls). Total extracted RNA was depleted of rRNA and treated with RNase R or water prior to high-throughput sequencing. The data were analyzed as shown in a schematic in Fig. 7A (see Materials and Methods). The procedure selects for stable circular RNA molecules but does not distinguish nuclear from cytoplasmic RNA. I will address the localization issue later, but can state at this point that many sisRNA are, in fact, in the cytoplasm.

In actively transcribing cells (untreated), I detected ~6000 RNase R resistant intronic RNAs from 65 nt to ~2000 nt in length. These transcripts derived from about one third of the introns from ~2,000 genes (Fig. 7B) with a bias for introns between 0.1 kb and 1.5 kb, a bias presumably due to RNA purification. About 50% of mapped branch points are adenine, the canonical branch point. Most of these intronic reads probably represent unstable lariats.

After a 12 hr-transcription inhibition, 493 ciRNAs were detected, evidence that 3T3 cells produce stable intronic RNAs. These RNAs were derived from 424 genes,

mostly from a single intron per gene (Fig. 7B). There was a strong bias for cytosine at the branch point (Fig. 7C). Notably, Dbr1 has reduced activity toward the C branch point (Nam *et al.*, 1994), suggesting that a C branch point may be important for sisRNA formation.

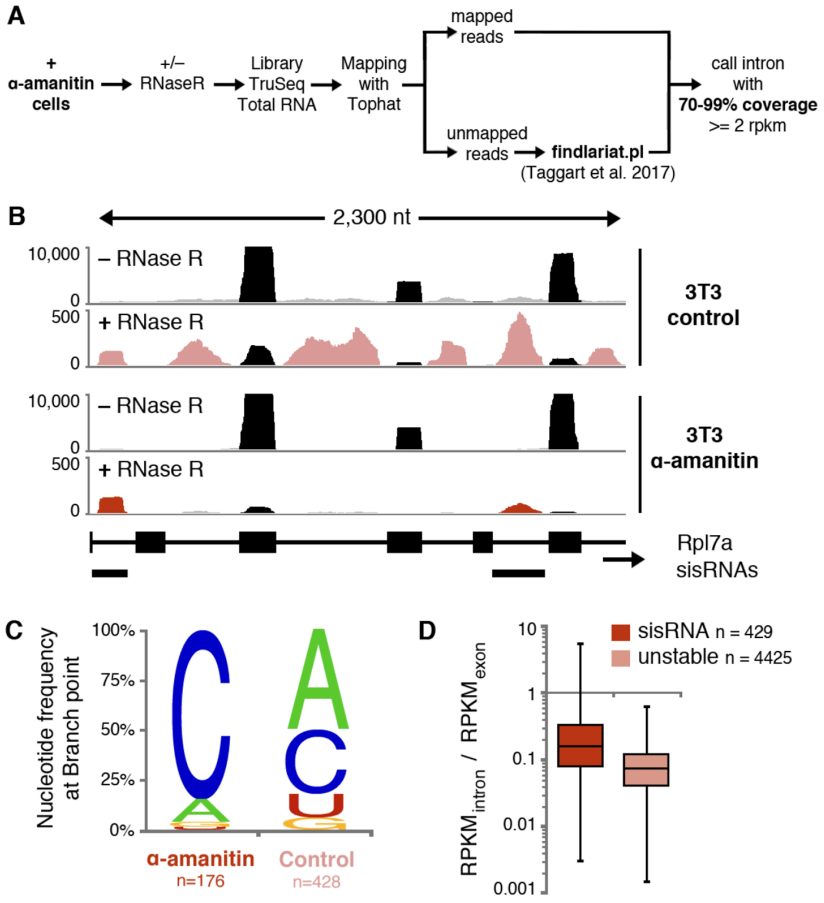


Figure 7. Identification of sisRNA in cell culture. (A) Schematic of the analysis performed to detect stable ciRNA. In bold are shown the crucial steps for identification of sisRNA. (B) IGV browser view of a typical gene for which we detect intronic RNAs in the control and the α-amanitin treated samples. The *upper* tracks (untreated cells) show that the Rpl7a mRNA is the dominant product of transcription (black) and that all intronic RNAs persist after RNase R treatment (pink). The *lower* tracks (α-amanitin treated cells) show the presence of the mRNA (black) and two intronic RNAs (red) after inhibiting transcription for 12 hrs. (C) Nucleotide frequency at the branch point of circular intronic RNA detected in the α-amanitin treated sample and control. 'n' represents the number of mapped branch points. (D) Abundance of intronic RNA relative to mRNA in untreated cells (n= number of intronic RNAs analyzed).

To further investigate the significance of the branch point, I compared two groups of circular intronic RNAs: (1) unusually *abundant* ciRNAs (> 5 rpkm) detected in control cells but not in α-amanitin treated cells and (2) *stable* ciRNAs detected in both control and treated cells. Both groups of ciRNA showed an average level of 10 rpkm but only the *abundant* ciRNAs tended to derive from highly expressed genes (average of 100 rpkm instead of 10 rpkm). The data suggest that the detection of the *abundant* ciRNAs is simply a reflection of high transcription rate and these ciRNAs represent transient lariats. I find that these two groups have distinct branch point signatures. In control samples, most *stable* lariats have a C branch point whereas most *abundant* ciRNA have an A branch point. Therefore, for equally abundant ciRNAs a C branch point is associated with stability in the 3T3 cell line.

Finally, because I detected both ciRNA and mRNA in rRNA depleted RNA from untreated cells, I can directly compare the relative abundance of an intronic RNA and its cognate mRNA. I find that about 70% of sisRNAs were at 10% of the mRNA levels or above, whereas ~70% of unstable intronic RNAs fell below the 10% mark (Fig. 7D).

In summary, I found that 3T3 cells produce stable circular intronic RNAs. These sisRNAs are characterized by (1) their short length, (2) their C branch point, and (3) their abundance relative to cognate mRNAs.

sisRNAs in different vertebrate species

To determine whether sisRNAs are found in other species, I extracted RNA from α-amanitin treated HeLa cells, DF1 cells (a chicken cell line) and XTC cells (a *Xenopus*

laevis cell line). For all samples, RNase R was used to enrich for circular RNA. I detected sisRNAs in all three species (Fig. 8A).

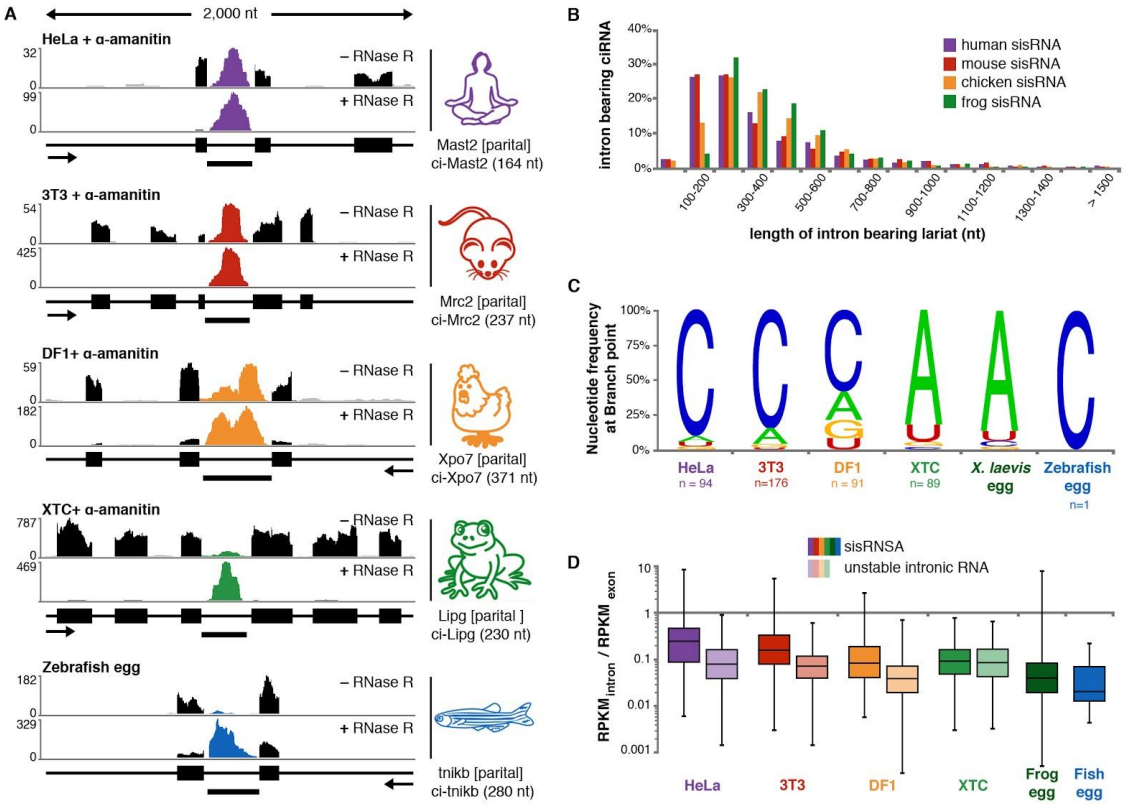


Figure 8. Identification of sisRNA in vertebrates. (A) A typical prominent sisRNA detected in human (purple coverage), mouse (red), chicken (orange), frog (green) and zebrafish (blue). RNase R was used to eliminate mRNA transcripts (black). Transcript annotation and orientation are shown below the coverage for each gene. (B) Number of sisRNAs relative to intron length (50-nt-bins). Human (purple), mouse (red), chicken (orange), frog (green) and zebrafish (blue) sisRNAs derive from short introns. (C) Nucleotide frequency at the branch point of sisRNAs (n = number of mapped branch points per species). (D) Abundance of 'stable' and 'unstable intronic RNAs relative to their cognate mRNA in untreated cells.

In HeLa cells, 421 sisRNAs derived from 354 genes, of which 68 also hosted a sisRNA in 3T3 cells. Similarity between human and mouse sisRNA includes the length of the sisRNAs (Fig. 8B), a bias for C branch point (Fig. 8C), and the ratio between the

sisRNA and its cognate mRNA level (Fig. 8D). In a previous study, Yang *et al*. (2014) analyzed 20 abundant lariats and determined that the branch point in HeLa cells could be either an A or a C nucleotide with about equal frequency. I confirmed 16 out of the 20 and mapped the branch point of 13. Among these 13 lariats, 8 lariats persisted after transcription inhibition, all of which had a C branch point. Conversely, I detected an A branch point in all five lariats that were detected in untreated cells only. This analysis is consistent with the interpretation that a C branch point is critical for sisRNA stabilization in mammals.

Similar observations were made on chicken DF1 cells. I detected 323 sisRNAs derived from short introns of 287 genes. These were stabilized as lariats with a cytosine branch point (55%), unlike unstable DF1 lariats with predominantly an adenine branch point (65%). Among the chicken sisRNA host genes, 21 were the same in HeLa cells and 32 in 3T3 cells. Overall, 6 genes were host to at least one sisRNA in humans, mice and chickens. The 6 common genes have very different functions, including transcription regulation, motors, ribosome biogenesis, fatty acid synthesis and protein localization.

In *Xenopus laevis* XTC cells, I found a somewhat different pattern: nearly 700 sisRNAs that derived from only 500 genes. In the most extreme case, one gene hosts 14 sisRNAs. I am sure of the stability, as many unstable transcripts are no longer detectable after inhibiting transcription for 12 hr. In addition, sisRNAs were barely detectable in untreated cells and most molecules had an A branch point. These sisRNAs from XTC cells resembled those I had observed in frog oocytes: thousands of sisRNAs, among which nearly 90% of mapped branchpoints were adenine (Fig. 8C).

To study sisRNA from a fish, I extracted RNA from freshly laid eggs of the zebrafish (*Danio rerio*). These eggs contain a large pool of stable RNA used during early embryogenesis. I treated egg RNA with RNase R to enrich for circular transcripts but found only 13 lariats. In Fig. 8 A, I show the read coverage for the *tnikb* gene and the lariat derived from its intron, the most abundant zebrafish sisRNA in our sample. This particular sisRNA has a cytosine branch point. Unfortunately, I could not map the branch point of the other sisRNAs because of their low abundance. An important limitation of our study is the dependence on genome annotation for sisRNA discovery. I anticipate that more sisRNA will be detected when a more complete genome assembly becomes available.

In summary, I found hundreds of circular sisRNA in six vertebrate species. These stable lariats derive from short introns and are characterized by a C branch point (except in *Xenopus*). They derive from coding and non-coding genes related to nearly all known cellular functions.

sisRNAs in mouse liver, brain, and blood

It has been advantageous to begin our study of sisRNA with cell lines which provide pure populations of cells in which transcription is easily inhibited. To establish whether sisRNAs occur in normal tissues, I extracted RNA from mouse liver, brain and whole blood and, as before, used RNase R to enrich for circular RNA before sequencing the samples. I found ~3,500 circular intronic RNAs in the liver, ~1,000 in the brain, and ~2,000 in whole blood. The coverage of the most abundant circular intronic RNAs in these tissue is shown in Fig. 9A.

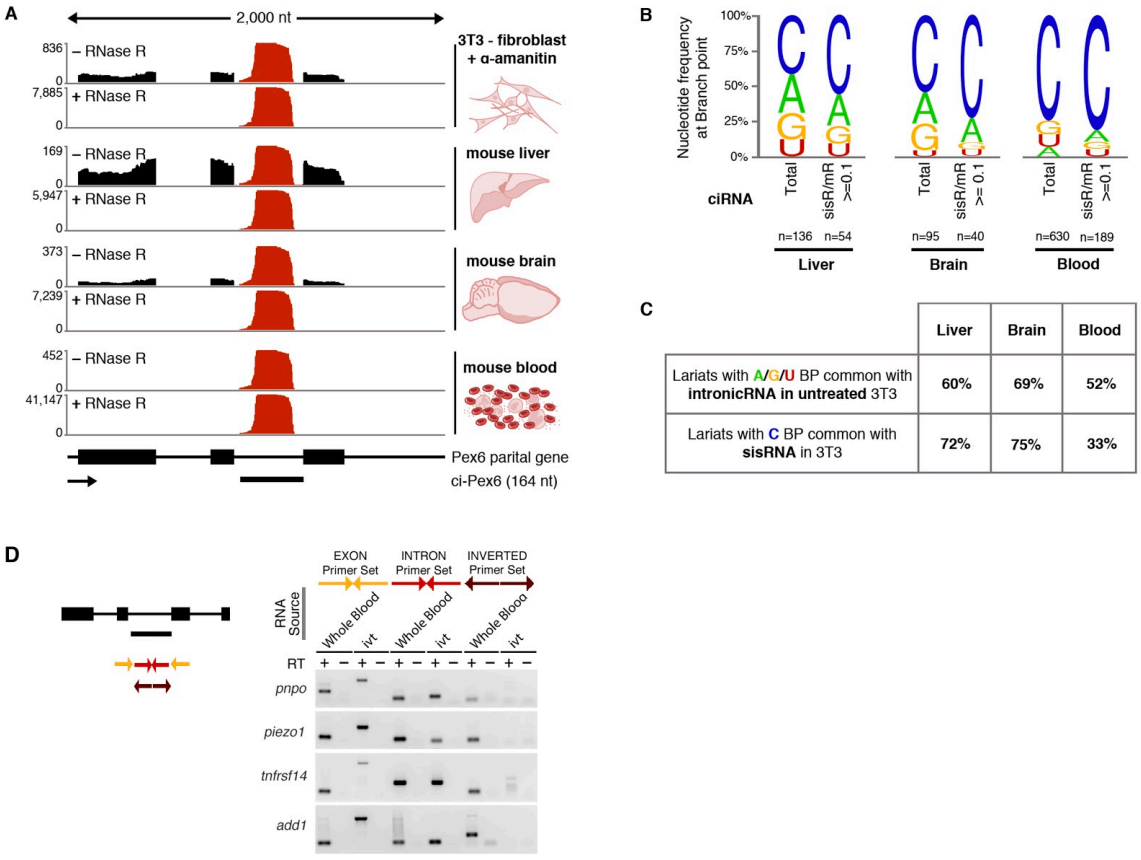


Figure 9. sisRNAs in different mouse tissues. (A) Read coverage from different tissues for mouse Pex6 gene. A very abundant sisRNA, ci-Pex6 (red), is detected in liver, brain, and whole blood. (B) Branch point nucleotide frequency of all lariats detected in liver, brain, and blood. 'n' represents the number of mapped branch points. Among lariats that are also found in α-amanitin treated 3T3 cells, 75% have a C branch point. (C) Percentage of C-branched vs. non C-branched lariats that are also detected in α-amanitin treated 3T3 cells. (D) Four whole blood sisRNA candidates were validated by RT-PCR. Primers were designed to amplify exonic sequences (yellow arrows), intronic sequences (red arrows), and circular intronic sequences (brown arrows facing outward). In vitro transcribed pre-mRNAs were used as controls.

Unlike RNA from α-amanitin-treated cultured cells, total RNA derived from whole tissues is a mixture of short lived and more stable molecules. To determine whether the circular intronic RNA in these tissues may be stable, I examined the three criteria met by sisRNAs in cell culture. I analyzed intronic RNA derived from short

introns (< 1500 nt) and determined the branch point usage. In both the liver and brain, roughly half of the mapped branch points were cytosine, suggesting the existence of sisRNA in these tissues (Fig. 9B). Next, I examined circular intronic RNAs that were at least at 10% of the mRNA level (as seen for most sisRNA detected in cell culture). I found 946 and 242 such intronic RNAs in the liver and brain respectively. These pools were enriched for C-branched lariats, consistent with their being sisRNAs (Fig. 9B). Taken together, the data suggest that we can detect both stable and unstable lariats. Importantly a fraction of these lariats has the characteristics of sisRNA: they are short, they have a C branchpoint, and they are expressed at 10% or more of their cognate mRNA level. These data provide the first evidence that sisRNA can accumulate in liver and brain.

To explore these sisRNA further, I divided them into two categories: lariats with a C branch point and those with A, G, or T. About 80% of C-branched lariats were detected in α-amanitin-treated 3T3 cells, confirming the presence of sisRNA in liver and brain. Conversely, about 80 % of the presumably unstable lariats (those with an A, G, or T branch point) were detected in untreated cells only (Fig. 9C). I was particularly interested in C-branched lariats and therefore I analyzed the ratio of these sisRNAs to their cognate mRNA. Surprisingly, I found that the ratio is relatively constant across tissues. In other words, the abundance of a particular sisRNA may be driven largely by its host gene's expression.

Finally, I searched for sisRNA in whole blood, a tissue with only a few cell types and with minimal transcription. As previously mentioned, I found about 2000 presumed circular intronic RNAs after RNase R enrichment. 75% of the mapped branch points were

cytosine, suggesting that most of these RNAs were sisRNAs (Fig. 9B). I performed RT-PCR on four candidates and validated our bioinformatic predictions (Fig. 9D).

Circular intronic RNA in red blood cells

Mammalian RBCs extrude their nuclei before going into circulation; hence, they are a source of pure cytoplasm. Because I had detected sisRNAs in whole blood, I asked whether these sisRNAs accumulated specifically in RBCs. I purified RNA from RBCs and subjected the sample to sequencing after rRNA depletion. Because of the low input, I did not treat the RNA with RNase R. I detected ~600 sisRNAs in mouse RBCs and ~300 sisRNAs in human RBCs. I show the read coverage of a typical human and mouse gene that encodes a sisRNA in Fig. 10A.

I am sure that the intronic RNAs correspond to circular sisRNA for several reasons. First, the lack of transcription in RBCs ensures that only stable transcripts were sequenced. Although I did not treat the RBC RNA with RNase R, I detected lariat reads, which predict the circularity of the transcripts. I confirmed the circular nature of a few candidates by RT-PCR and Northern blotting (Fig. 10B and C). Finally, I found that the vast majority of the lariats had a C branch point (Fig. 10D) and were much more abundant than their cognate mRNA (when the mRNA was detectable).

Because most cognate mRNAs showed little to no reads, one possibility was that the sisRNAs did not originate from the RBCs themselves. To test this possibility, I analyzed publicly available data of mouse and human erythroblasts, the nucleated RBC precursors in the bone (An *et al.*, 2014). For each sisRNA detected in mouse and human RBCs, I found that the cognate mRNA was expressed in erythroblasts (Fig. 10A). Hence,

during RBC maturation, it is likely that mRNAs were degraded while lariats persisted. An important conclusion is that sisRNAs can occur in the cytoplasm of mammalian RBCs, which lack both transcription and translation machineries.

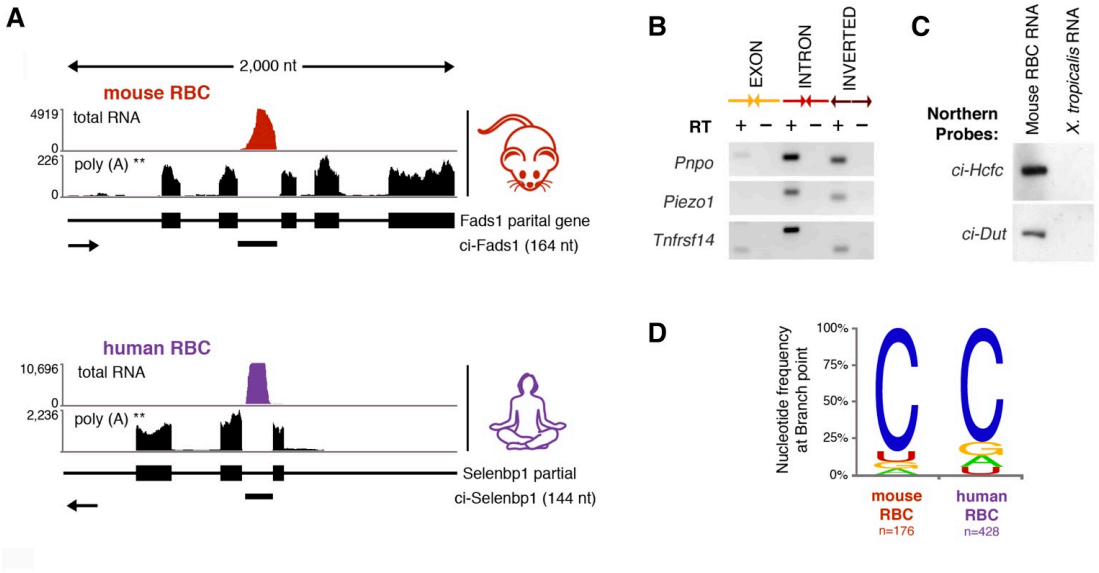


Figure 10. sisRNA in mammalian red blood cells. (A) reads of a typical sisRNA found in mouse and human RBCs. The *upper* track corresponds to total RNA from mature RBCs and the lower track corresponds to public RNAseq data of polyadenylated transcripts from RBC precursor cells (An *et al*., 2014). (B) RT-PCR using exonic primers (yellow) and intronic primers (red). RBC Intronic RNAs can be amplified with inward and outward primers (brown), confirming that these RNAs are circular. (C) Nucleotide frequency at the branch point.

In situ hybridization detects intronic RNA in cytoplasm

Definitive evidence that stable lariat can be stored in the cytoplasm in vertebrate cells (other than RBCs) comes from single molecule *in situ* hybridization carried out by Dr. Joseph Gall. We chose probes for three highly abundant sisRNAs, one each from human, mouse and chicken (Fig. 11A). The human and mouse sisRNAs were known to be abundant in RBCs as well as in cultured cells. Because all of these introns are less than

250 nucleotides in length (Fig. 11A) we used probes consisting of 3-4 pairs of short oligonucleotides (designed by Advanced Cell Diagnostics, Inc.), a multi-step signal amplification series, and detection by Fast Red. Positive (exonic *PpiB*) and negative (bacterial *DapB*) controls were included. In each case we detected weak but clear positive signals in the cytoplasm corresponding to the intronic probes (Fig. 11B). Because the tissues had not been treated with α-amanitin to inhibit ongoing transcription, we also saw positive signals in the nucleus that probably correspond to nascent transcripts at the sites of transcription and to unstable spliced introns.

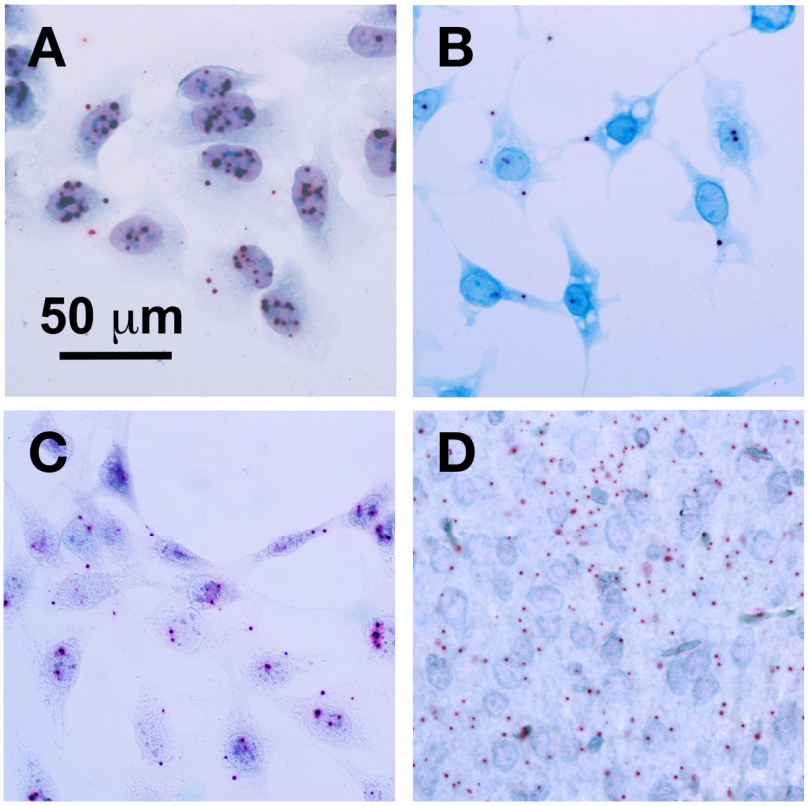


Figure 11. Single Molecule In Situ Hybridization of Introns carried by Dr. Joseph Gall. (A) Human Hs ANAPC2 intron 11 in cultured human cells (HeLa). (B) Chicken Gg ECE1intron 17 in cultured chicken cells (DF1). (C) Mouse Mm Vars intron 25 in cultured mouse cells (3T3). (D) Mouse Mm Vars intron 25 in section of mouse hind brain.

sisRNAs are exported from the nucleus by NXF1/NXT1

Because the oocyte nuclear envelope remains intact during the entire course of oogenesis, it seemed probable that sisRNAs reach the cytoplasm by an active transport mechanism. To test this possibility, I injected various constructs into the oocyte nucleus and monitored the nuclear and cytoplasmic distribution of their products. It is well-known that the oocyte stores large amounts of the proteins required for all export machineries (Pante, 2006). Also, because export is a saturable process, one can perform export competition assays by injecting an RNA of interest along with large amounts of competitor RNA. After a few minutes, the oocytes can be dissected and cytoplasmic accumulation can be determined (Dargemont and Kuhn, 1992; Jarmolowski *et al*., 1994). I began by injecting purified sisRNA or pre-mRNA into the nucleus. However, these RNAs were rapidly degraded, suggesting that sisRNAs are normally stabilized at or near the time of transcription. Hence, to study the export mechanism, I expressed sisRNAs from injected DNA plasmids.

In our system, I express ectopic sisRNAs independently of their endogenous cognate mRNA. *X. tropicalis* introns were inserted into a partial *X. laevis Ncl* gene between exons 14 and 15, a combination that shows high splicing efficiency. The partial Ncl gene was placed downstream of an mCherry gene. Once injected into the nucleus, the DNA construct is transcribed, giving rise to both *X. tropicalis* sisRNA and mCherry mRNA (Fig. 12C). In this way, I can monitor the efficiency of injection, transcription, and translation by selecting for red oocytes (Fig. 12A and B). Because intronic sequences

are highly divergent between *X. tropicalis* and *X. laevis*, we can easily distinguish ectopic from endogenous sisRNAs.

Two days after injection, I examined the amount of sisRNA that had accumulated in the cytoplasm (Fig. 12D). For one of the most abundant *X. tropicalis* cytoplasmic sisRNAs, ci-faf2, 50% of the newly transcribed lariats were detected in the cytoplasm. One the other hand, for ci-actg1, a sisRNA found primarily in the nucleus, only 5% of the ectopic transcripts accumulated in the cytoplasm. To determine which export system is used by ciRNA, I co-injected constructs whose transcripts use three different export pathways: an mRNA (GFP), a *Xenopus* tRNA, and a short hairpin RNA (against a mouse transcript not conserved in *Xenopus*). The export of sisRNA was inhibited only by overexpression of GFP mRNA, suggesting that mRNA and sisRNA share a common export pathway. Importantly, the fact that sisRNA export can be inhibited in this way strongly implies that sisRNA export is not due to simple leakage from the nucleus.

Most mRNAs are exported via NXF1/NXT1 (reviewed by Delaleau and Borden, 2005). To test whether NXF1 is sufficient for sisRNA export, I co-injected a ci-faf2 DNA construct along with increasing concentrations of GFP with and without NXF1 DNA. As expected, increasing amounts of GFP mRNA led to a decrease in sisRNA export (Fig. 12E, green data points, semi-quantified in Fig. 12F), whereas co-injection of an NXF1 DNA facilitated the export of sisRNA in the presence of the competitor GFP mRNA (Fig. 12E, black data points). From these experiments, I conclude that sisRNAs are selectively exported to the cytoplasm using the NXF1 machinery.

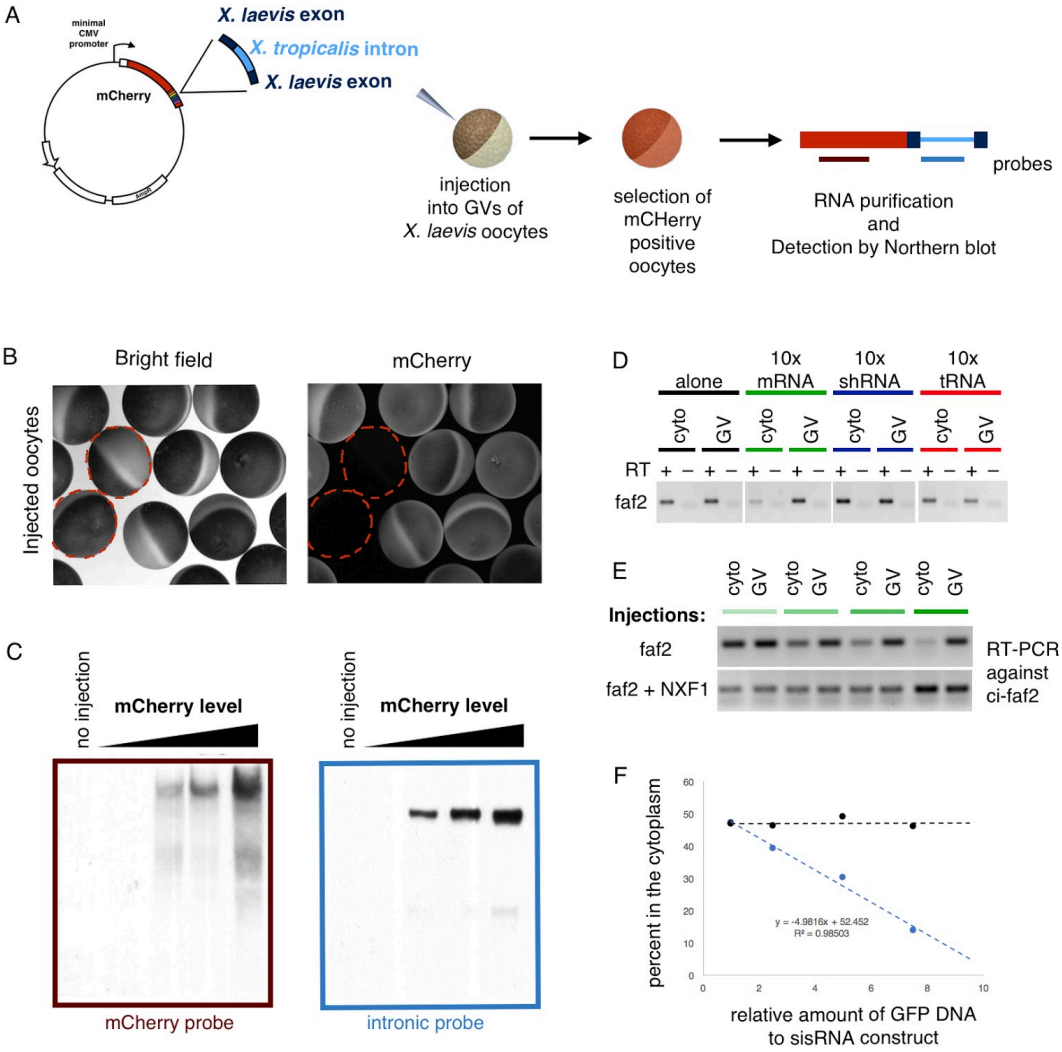


Figure 12. NXF1 exports sisRNA to the cytoplasm. (A) Diagram of the method and construct used to express *Xenopus tropicalis* sisRNAs in *Xenopus laevis* oocytes. (B) mCherry can be detected under a fluorescent dissecting microscope. (C) Both mCherry mRNA and faf2 sisRNA are detected by northern blot 2-days post injection. Brighter oocytes contained more sisRNA. (D) RT-PCR against ci-faf2 on cytoplasmic and nuclear fractions. Oocytes were injected with faf2 intron construct alone (black) or with GFP DNA (mRNA, green), an shRNA construct (blue) and a tRNA construct (red). (E) Increasing amount of co-injected GFP DNA correlates with decrease of cytoplasmic sisRNA in the cytoplasm. Co-injection with NXF1 rescues the cytoplasmic accumulation of ci-faf2. (F) Semi-quantification of the RT-PCR shown in E.

Cytoplasmic sisRNAs are not associated with their cognate mRNAs

A possible function for cytoplasmic sisRNAs would be to regulate translation or localization of their cognate mRNAs. To explore this possibility, we asked whether sisRNAs might be physically associated with their cognate mRNAs in the cytoplasm. We carried out in situ hybridization of the products of two genes in the cytoplasm of *Xenopus tropicalis* oocytes (uggt1 = UDP-glucose glycoprotein glucosyltransferase 1 and pphln1 = periphilin 1). These two genes each produce a very long sisRNA, making it possible to design probes consisting of multiple fluorescent oligonucleotides for maximal signal intensity. Probes were labeled with Quasar 570 (mRNA) or Quasar 670 (intron) and hybridized individually and simultaneously. Very small transparent oocytes were chosen so that the probes could be visualized in whole mounts. In situ hybridization of probes against uggt1 mRNA and its longest sisRNA (Fig. 13A and B) show signal in the cytoplasm and essentially no colocalization (Fig. 13C). Because *Xenopus tropicalis* and *Xenopus laevis* show high conservation of mRNA but not intronic sequence, we probed for *Xenopus tropicalis* uggt1 mRNA and sisRNA in *Xenopus laevis* oocytes. Reassuringly, we observed signal only with exonic probes (Fig. 13D). Similar results were obtained for *Xenopus tropicalis* pphin1 mRNA and its longest intron. Although the mRNA and cognate sisRNA from each of these two genes might transiently associate in the cytoplasm, they do not form a permanent association detectable by in situ hybridization.

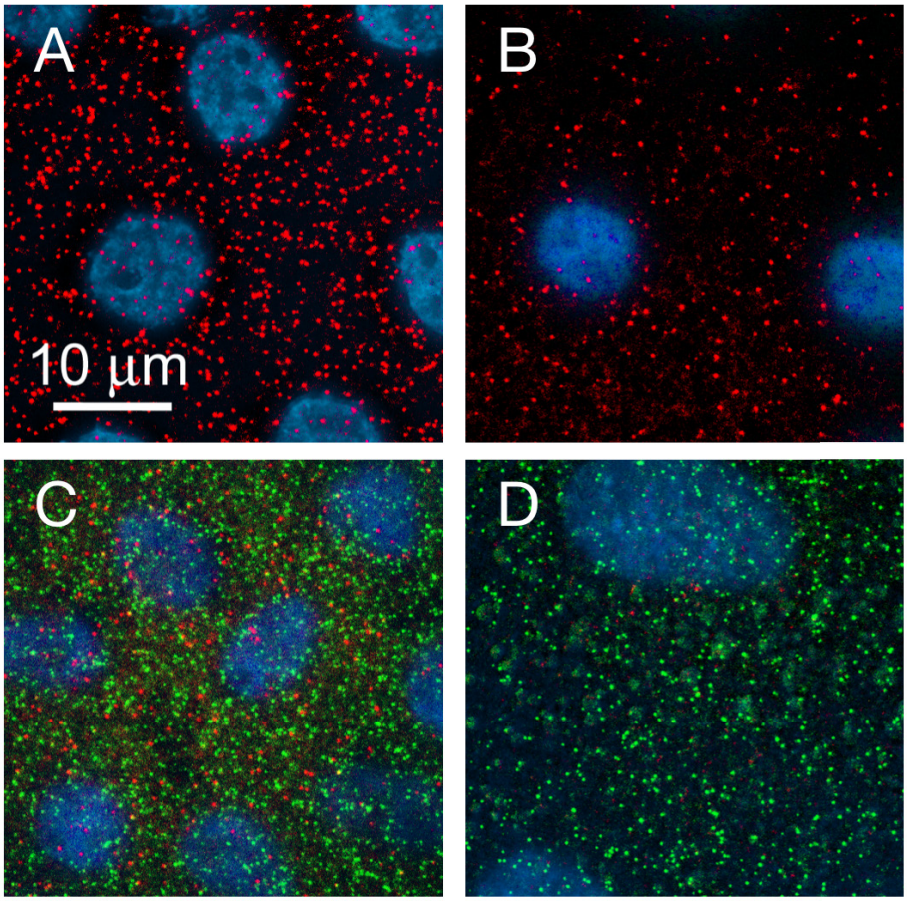


Figure 13. **Single molecule in situ hybridization of exon and intron probes from the Xenopus tropicalis uggt-1 gene carried out by Dr. Joseph Gall.** Oocytes (200-400 µm diameter) were hybridized and examined as whole mounts. Images are deconvolved stacks extending from the follicle cells (blue nuclei, DAPI stain) surrounding the oocyte into the peripheral cytoplasm of the oocyte. All of the signal is in the oocyte itself. (A) *X. tropicalis* oocyte hybridized against the uggt-1 exon probe (red). (B) *X. tropicalis* oocyte hybridized against the uggt-1 intron probe (red). (C) *X. tropicalis* oocyte hybridized against both the exon probe (green) and the intron probe (red). The two probes give non-overlapping signals. (D) *X. laevis* oocyte hybridized with both probes, as in C. Only the green exonic probe gives a positive signal in this species.

DISCUSSION

Lariat RNA in the cytoplasm in all vertebrates

In this study, I show that circular sisRNAs exist in human, mouse, chicken, frog, and fish. By inhibiting transcription for 12 hrs, I distinguished sisRNA from the abundant unstable intronic RNAs and found many circular sisRNAs. Host genes were typically housekeeping genes, but there was little overlap across species. Importantly, we detected many circular sisRNAs in the cytoplasm of the vertebrate cells, ruling out the possibility that cytoplasmic sisRNA are restricted to *Xenopus* oocytes.

In a previous study, Zhang et al. (2013) described the existence of nuclear circular intronic RNAs. They ruled out a cytoplasmic localization by carrying out an RT-PCR analysis against certain ciRNA on nuclear and cytoplasmic RNA, and they performed *in situ* hybridization experiments against one of their candidates. Notably, their conclusion conflicts with our observation; however, in their studies, both experiments were carried out on actively transcribing cells, thus providing no distinction between stable and unstable intronic RNA. Additionally, we had observed that FISH experiments do not resolve cytoplasmic sisRNAs, which would explain why they did not detect ciRNAs in the cytoplasm. Instead, because cytoplasmic sisRNAs do not appear to aggregate in cells, their signal must be resolved by single molecule in situ hybridization, as we have demonstrated in this chapter. I argue that the existence of sisRNAs in the cytoplasm is irrefutable because we detect sisRNAs in RBCs which lack nuclei, guarantying the cytoplasmic localization of these RNAs.

Mechanism of stability

By comparing stable and unstable lariats, I found that stable lariats have a C branch point instead of the canonical A branch point (except in *Xenopus*). As mentioned, the debranching enzyme is relatively inefficient in linearizing C branched lariats. I speculate that the delay in linearization 'permits' the formation of a stable complex for several reasons. It is worth noting that about 10% of circular sisRNAs in mammals were A-branched and not all C-branched lariats were stabilized. Therefore, other features, such as structure, must also trigger a delay in linearization.

In *Xenopus* cell culture and oocytes, circular sisRNAs are mostly A-branched. Intriguingly, *Xenopus* sisRNAs are ten times less abundant than mammalian sisRNAs on average. Nevertheless, there are 600 different circular intronic RNAs stabilized in *Xenopus* XTC cells, while only 300 to 400 sisRNAs accumulate in 3T3 and HeLa cells. Ultimately, *Xenopus* cells may stabilize roughly similar numbers of circular intronic molecules as do mammalian cells.

sisRNA may function independently of their cognate mRNA

Because sisRNAs were first found in the nucleus, it has been proposed that these RNAs regulate transcription of their host genes (Zhang *et al.*, 2013; Pek *et al.*, 2015; Tay and Pek, 2017). However, here we presented evidence suggesting that not all sisRNAs regulate their host genes. (1) Some sisRNAs accumulate in the cytoplasm instead of the nucleus. (2) Host genes are different across species. Here I show that human and mouse RBCs have little overlap between their sisRNA host genes. (3) sisRNA is found in RBCs, which lack transcription and translation machinery and have little mRNA. (4) In *Xenopus*

oocytes, we show that sisRNAs do not co-localize with their cognate mRNA. Altogether, these data suggest that sisRNAs can function independent of their cognate mRNA.

It is worth noting that failing to recycle intronic RNAs can be detrimental to cells. In fact, knockout of the debranching enzyme (Dbr1) can be lethal, and down-regulation of Dbr1 can contribute to tumor formation (Han et al., 2017). Therefore, the fact that some intronic RNAs escape in 'normal' tissues cannot be random. We think it is most likely that sisRNAs are important 'sponges' or regulators for RNA binding proteins (see Chapter 3 for a specific example). Consistent with this speculation, when downregulation of Dbr1 is induced, intronic lariat RNAs accumulate and associate with RNA binding proteins (RBPs) such as TDP-43, Dicer or snRNPs (Armakola et al., 2012, Li et al., 2016 and Han et al., 2017). Perhaps circular sisRNAs bind RBPs to lessen aggregation or availability of the proteins. In that case, stable lariats would be essential for normal homeostasis of cells.

MATERIALS AND METHODS

Cell culture

HeLa and 3T3 cells were cultured at 37°C in DMEM (Gibco) with 10% fetal bovine serum (FBS). DF1 cells were cultured at 39°C in DMEM with 10% FBS. XTC cells (kindly provided by Helene Cousin) were cultured at room temperature in 66% RPM medium with 10% FBS.

Tissue collection

Blood was collected by Safia Malki from asphyxiated C/D1 or Bl/6 mice by cardiac puncture with an EDTA-coated 1 ml syringe and a 22-gauge needle. Brain and liver were dissected from the same animals and immediately processed for RNA extraction.

Human red blood cells were purchased from the Interstate Blood Bank, Inc (via Zen-Bio company). Prior to shipment, the sample tested negative for hepatitis B, hepatitis C, human immunodeficiency virus and syphilis. The samples were stored at between 2°C and 8°C in anticoagulant citrate dextrose solution.

Red blood cells purification

Blood samples were centrifuged at 300 g for 10 min at room temperature and the plasma was removed. Cells were resuspended in 1X PBS, layered on a 50%-80% Percoll gradient and centrifuged at 1,000 g for 30 min at room temperature. RBCs settled near the bottom of the tube and were collected free of white blood cells. Before usage, RBCs were washed 3 times in 1X PBS.

RNA purification

RNA was extracted with TRIzol reagent (Ambion) and purified with the Direct-zol kit (Zymo Research). DNase treatment was performed on the Direct-zol columns according to the manufacturer's protocol. RNA was quantitated with a Nanodrop 2000 spectrophotometer (Thermo Scientific) and assessed with a Bioanalyzer 2100 (Agilent).

RNase R

RNase R treatment was carried as described in Chapter 2. In short, RNA was denatured 5 min at 72℃ in RNase R 1X buffer and 1 µL of RNase R (Epicenter) was added to the reaction prior to overnight incubation at 37℃. RNA degradation was prevented by adding 1 µL RNasin (Promega).

Library preparation, sequencing, and sequence analysis

Libraries were prepared using TrueSeq stranded total RNA sample preparation (Illumina). Sequencing was performed on either an Illumina HiSeq 2000 sequencer with 100 bp single-end reads or with 150 bp single-end reads. Reads were aligned with TopHat (v2.0.7) to the mouse genome (v10), the human genome (v19), the chicken genome (v5), or the *X. laevis* genome (v9.0). Intronic reads were quantified using Bedtools (v2.15.0) and lariat reads were obtained using lariat_find.pl (Taggart *et al*., 2017). Specific cut offs used are detailed in the schematic in Fig. 1A.

RT–PCR analysis

To confirm the bioinformatic predictions, I performed RT-PCR analysis according to a previously published protocol (Talhoaurne and Gall, 2014). In short,

cDNA was made using Episcript (Epicenter) and amplified with Taq (Qiagen) and various primers (sequence provided in supplementary Table S2, Appendix). In vitro transcribed RNA was used as a control to demonstrate the efficiency of the primers used.

Northern blots

Up to 5 μg of RNA per sample was separated on an 8% polyacrylamide–8 M urea gel and transferred onto a nylon membrane (Zeta Probe GT, Bio-Rad). RNA was probed with a digoxigenin (DIG)-labelled PCR product in hybridization buffer (Roche) at 42℃ and detected using an anti-DIG antibody conjugated with alkaline phosphatase and CDP-*Star* chemiluminescent substrate (Roche).

Single molecule in situ hybridization

Experiments were carried by Dr. Joseph G. Gall according to the manufacturer's protocols (RNAScope and BaseScope probes from ACDBio and Stellaris probes from LGC Biosearch Technologies).

Xenopus oocytes and injections

Pieces of ovary were surgically removed from anesthetized *X. laevis* females (MS222, pH 7) and cultured in OR2 medium at 16° (Wallace *et al*., 1973). Manually separated oocytes were microinjected at the animal pole with 100 pg of plasmid in a 2.4 nl volume using the Nanoject II microinjection apparatus (Drummond Scientific). Needles were pulled from capillary tubes (0.5-mm inner diameter, 1.2-mm outer diameter) with a horizontal micropette puller (model P-97, SUTTER Instrument). Different amounts of competitor plasmid were co-injected: 100, 250, 500, 750 and up to 1000 pg. Injecting more than 1000 pg of plasmids was often lethal for oocytes. 48 hr after

injections, oocytes were dissected in OR2 isolation solution at pH5.6 (83 mM KCl, 17 mM NaCl, 6.0 mM Na_2HPO_4, 4.0 mM KH_2PO_4, 1 mM $MgCl_2$, 1.0 mM dithiothreitol). GVs and cytoplasms were collected on dry ice before adding TRIzol for RNA extraction.

Plasmids

A construct that expresses sisRNAs was generated as follow. The mCherry-CAAX gene (kindly provided by the Halpern laboratory) was subcloned into the pGMT plasmid. Multicloning sites were used to add a partial *Xenopus laevis* gene (exon 14 to 15 of *Ncl*) in which the intron was replaced by intron 2 of the *Xenopus tropicalis faf2* gene. The plasmid encoding HA-tagged (N terminus) human NXF1 was purchased from ABclonal. The pCS2 GFP plasmid was generated in the Donald Brown laboratory. A *Xenopus tropicalis* tRNA gene was amplified by PCR and cloned in the pGMT plasmid. An shRNA construct was kindly provided by the Zhang laboratory.

REFERENCES

An X. Schulz V.P. Li J. Wu K. Liu J. Xue F. Hu J. Mohandas N. Gallagher P.G. **2014**. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* 123(22), 3466-3477.

Armakola M, Higgins MJ, Figley MD, Barmada SJ, Scarborough EA, Diaz Z, Fang X, Shorter J, Krogan NJ, Finkbeiner S, *et al*. **2012**. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. Nat Genet 44: 1302–1309

Berget SM, Moore C, Sharp PA. **1977**. Spliced segments at the 59 terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci 74: 3171–3175.

Chorev M and Carmel L. **2012**. The Function of Introns. Front Genet. 3: 55.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. **1977**. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12: 1–8.

Dargemont C, Kuhn LC. **1992**. Export of mRNA from microinjected nuclei of *Xenopus laevis* oocytes. J Cell Biol. 118:1–9.

Delaleau M and Borden K. **2005**. Multiple Export Mechanisms for mRNAs. Cells. *4*(3), 452-473

Han B, Park HK, Ching T, Panneerselvam J, Wang H, Shen Y, Zhang J, Li L, Che R, Garmire L, Fei P. **2017**. Human DBR1 modulates the recycling of snRNPs to affect alternative RNA splicing and contributes to the suppression of cancer development. Oncogene. 36(38):5382-5391.

Jarmolowski A, Boelens WC, Izaurralde E, Mattaj IW. **1994**.Nuclear export of different classes of RNA is mediated by specific factors. J. Cell Biol., 124 , pp. 627–635

Kimura T, Hashimoto I, Nishikawa M, Yamada H. **2009**. Nucleocytoplasmic transport of luciferase gene mRNA requires CRM1/Exportin1 and RanGTPase. Med Mol Morphol. 42(2):70–81.

Gardner EJ, Nizami ZF, Talbot CC Jr, Gall JG **2012**. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. Genes Dev 26: 2550–2559

Li Z, Wang S, Cheng J, Su C, Zhong S, Liu Q *et al*. **2016**. Intron Lariat RNA Inhibits MicroRNA Biogenesis by Sequestering the Dicing Complex in Arabidopsis. PLoS Genet 12: e1006422

Mercer TR, Dinger ME, Sunkin SM, Mehler MF and Mattick JS. **2008**. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A*. 105(2): 716–721.

Maquat LE. **2004**. Nonsense mediated mRNA decay: splicing, translation and mRNP dynamics. Nat. Rev. Mol. Cell. Biol. 5, 89–99.

Moore MJ. **2005**. From birth to death: the complex lives of eukaryotic mRNAs. Science 309, 1514–1518

Nakaya HI, Amaral PP, Louro R, Fachel AA., Moreira YB, Tarik A El-Jundi TA, da Silva AM, Reis EM and Verjovski-Almeida S. **2007**. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biology*. 8(3):R43.

Nam K., Hudson R.H., Chapman K.B., Ganeshan K., Damha M.J., Boeke J.D. **1994**. Yeast lariat debranching enzyme. Substrate and sequence specificity. *J. Biol. Chem*. 269:20613–20621.

Ooi SL, Samarsky DA, Fournier MJ, Boeke JD **1998**. Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: Intron length effects and activity of a precursor snoRNA. RNA 4: 1096–1110

Pante N. **2006**. "Use of intact *Xenopus* oocytes in nucleocytoplasmic transport studies," in *Xenopus Protocols: Cell Biology and Signal Transdution* ed. Liu J., editor. (Totowa, NJ: Humana Press Inc.) 301–314.

Pimentel H, Parra M, Gee SL, Mohandas N, Pachter L, Conboy JG. **2016**. A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Research*. 44(2):838-851.

Sullivan JC, Reitzel AM, Finnerty JR. **2006**. A high percentage of introns in human genes were present early in animal evolution: evidence from the basal metazoan Nematostella vectensis. *Genome Inform*.17(1):219-29.

Taggart AJ, Lin C-L, Shrestha B, Heintzelman C, Kim S and Fairbrother WG. **2017**. Large-scale analysis of branchpoint usage across species and cell lines. Genome Res. 27, 639–649.

Talhouarne GJS and Gall JG. **2014**. Lariat intronic RNAs in the cytoplasm of Xenopus tropicalis oocytes. RNA. 20:1476–87.

Tay MLI, Pek JW. **2017**. Maternally inherited stable intronic sequence RNA triggers a self-reinforcing feedback loop during development. *Current Biology* 27 (7), 1062-1067

Vasil, V., Clancy, M., Ferl, R. J., Vasil, I. K., and Hannah, L. C. **1989**. Increased gene expression by the first intron of maize shrunken-1 locus in grass species. Plant Physiol. 91, 1575–1579.

Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG and Chen LL. **2012**. Long noncoding RNAs with snoRNA ends. *Mol Cell*. 48:219-30.

Wallace RA, Jared DW, Dumont JN, Sega MW. **1973**. Protein incorporation by isolated amphibian oocytes: III. Optimum incubation conditions. J Exp Zool 184: 321-333.

Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL **2013**. Circular intronic long noncoding RNAs. *Mol Cell* 51: 792–806

CHAPTER 3

Stable lariats bearing a snoRNA in vertebrate cells

INTRODUCTION

In previous studies (Chapters 1 and 2), I established the existence of stable intronic sequence (sis) RNAs in both the nucleus and cytoplasm. The most abundant sisRNAs are stabilized in the form of lariats. That is to say, unlike typical intronic RNAs, sisRNAs can escape linearization after splicing, which prevents further processing. In higher eukaryotes, many introns encode ncRNAs such as miRNA and small nucleolar (sno)RNA.

The majority of snoRNAs are guides for site-specific uridine isomerization (H/ACA box snoRNAs) and 2'-O-methylation (C/D box snoRNAs) of ribosomal RNA (rRNA), small nuclear RNA (snRNA), and potentially other RNAs (reviewed by Yu *et al*., 2005). In vertebrates, almost all snoRNAs are encoded by introns. Therefore, snoRNAs are first released in the form of a lariat, which must be linearized by the debranching enzyme, Dbr1, before maturation. Alternatively, lariats may be linearized by endonucleases to promote snoRNA maturation (Petfalski *et al*., 1998).

In this study, I searched for snoRNA motifs in sisRNAs and found a subset of sisRNAs that overlapped with annotated snoRNA or snoRNA-like motifs. These stable lariats accumulate in the nucleus *and* cytoplasm. We discuss their potential functions.

RESULTS

A stable lariat bearing a snoRNA (slb-snoRNA) in vertebrates

In my previous study, I identified sisRNAs in human, mouse, chicken, and frog (Chapter 2). For this study, I compared my list of sisRNAs with annotated snoRNAs. Interestingly, I found that 30 sisRNAs that overlapped with an annotated snoRNA (Fig. 14A). To extend the search to unannotated snoRNAs, I used the snoRNA motif prediction software snoReport 2.0 (De Araujo Oliveira *et al*. 2016). Across all four species analyzed, 182 sisRNAs were predicted to have a snoRNA motif, among which 28 correspond to the annotated snoRNAs mentioned above, and 85 others showed high predictability scores (Table S1).

We know that these particular sisRNAs are circular by three criteria. (1) The intronic reads mapped beyond the snoRNA sequence and extended from the 5'end to the branch point. (2) The exonuclease RNase R failed to degrade these sisRNAs. Finally, (3) I detected reads corresponding to the 5'-branch point junction of the lariat (Fig. 14A, 'Inverted reads' tracks). Importantly, I confirmed the circular nature of the intronic RNA by RT-PCR. First, I generated cDNA with RNA purified from α-amanitin treated cells. I used either Episcript, a reverse transcriptase that can go across branch points, or AMV-RT, a reverse transcriptase which terminates at a branch point (Talhouarne and Gall, 2014). Intronic cDNA produced with either enzyme could be amplified by PCR using inward facing primers. However, with outward facing primers, only cDNA generated with Episcript was amplified (Fig. 14B). The specificity of the PCR products was confirmed by sequencing. I concluded that lariats bearing a snoRNA could be stabilized

in vertebrates. I refer to these circular transcripts as stable lariat bearing a snoRNA or slb-snoRNA.

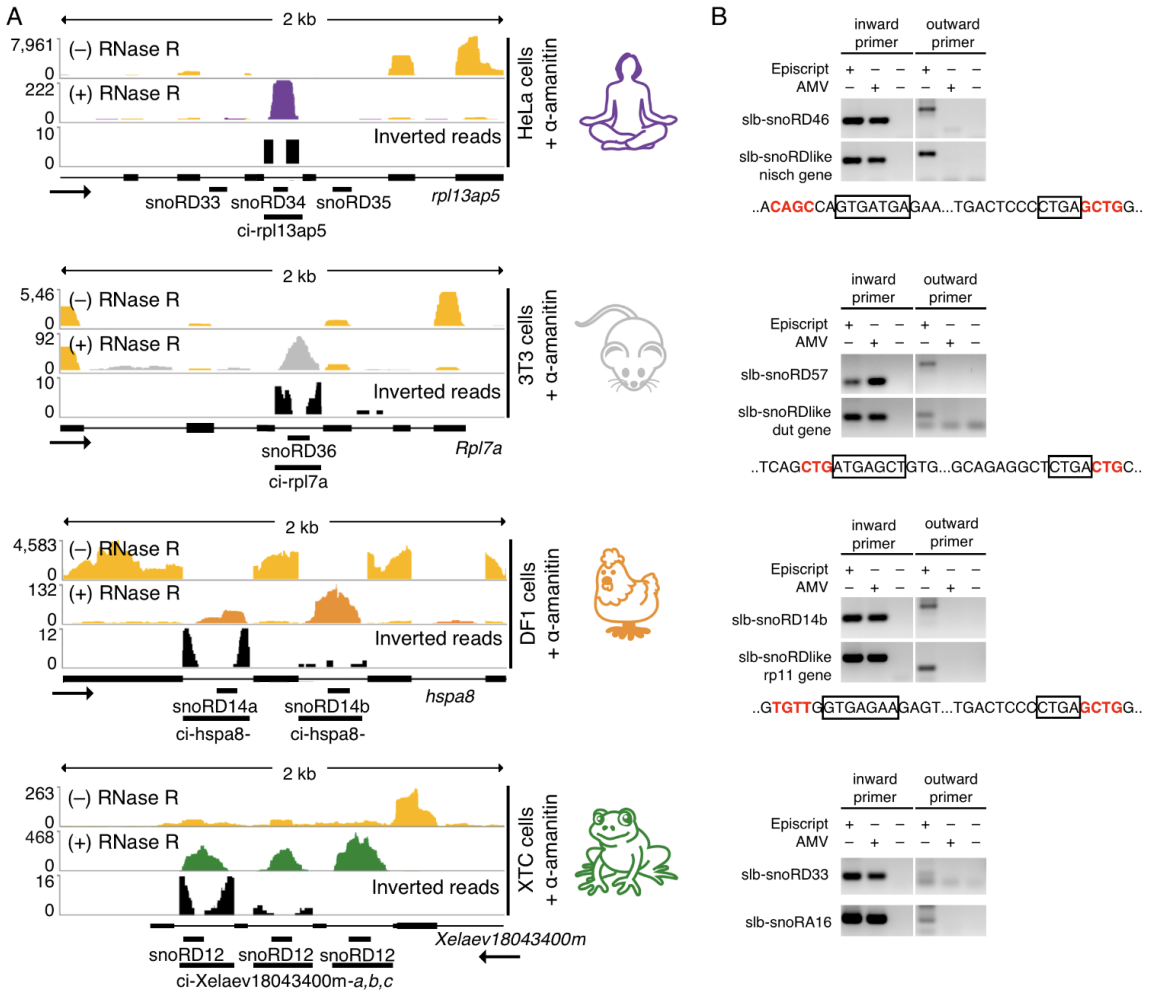


Figure 14. Stable lariats bearing a snoRNA (slb-snoRNA) are detected in human, mouse, chicken and frog. (A) IGV browser view of typical genes encoding a stable circular intron that overlaps with a snoRNA. Unlike exonic reads (yellow), intronic reads are detected in the RNase R samples, shown in purple (human), grey (mouse), orange (chicken) and green (frog). In black is the coverage of intronic 'inverted' reads. These reads demonstrate that the intronic transcripts are circular. (B) RT-PCR experiments for 2 slb-snoRNA candidates from human, mouse, chicken and frog. Reverse transcription was carried out with either Episcript or AMV-RT. Intronic cDNA was amplified by PCR carried out with inward and outward primers (to detect circular molecules only). Below the RT-PCR experiments, predicted C/D box motifs for the tested slb-snoRDlike transcripts are shown. Red sequences form the stem motif and the boxed sequence represent boxes C and D.

slb-snoRNAs can accumulate in the cytoplasm

In a previous study, I searched for cytoplasmic sisRNAs in mammalian red blood cells (RBCs). These cells lack nuclei and therefore provide a pure cytoplasmic fraction. I found hundreds of lariats in the RBC RNA samples. To my surprise, two sisRNAs overlapped with annotated snoRNAs (Fig. 15A). The detection of human slb-snoRD34 and mouse slb-snoRD66 in RBCs is proof that slb-snoRNAs can accumulate in the cytoplasm. To further analyze cytoplasmic slb-snoRNA, I turned to the *Xenopus* oocyte for its numerous advantages. First, the *Xenopus* oocyte makes possible the collection of pure nuclear and cytoplasmic fractions. Furthermore, *Xenopus* oocytes produce thousands of nuclear and cytoplasmic sisRNAs (Talhouarne and Gall 2014). Finally, we recently annotated nearly all functional snoRNAs in *Xenopus tropicalis* (Deryusheva, Talhouarne and Gall, in preparation).

First, I analyzed the RNAseq data generated from GV RNA and found that the vast majority of the annotated snoRNAs were detected both in the mature form (green coverage in Fig. 15B, top track) and in stable lariats (red coverage in Fig. 15B, top 2 tracks). I confirmed the bioinformatic analysis by northern blot against the *Xenopus tropicalis* snoRD69 sequence. I detected a band about 70 nucleotides in length, corresponding to the mature form, and a higher band corresponding to the stable lariat. I find that slb-snoRND69 is less abundant than snoRD69, consistent with the RNAseq data. Moreover, about a quarter of all nuclear sisRNAs were predicted to have snoRNA motifs.

More surprisingly, 16 abundant cytoplasmic sisRNAs contain snoRNAs (blue coverage in Fig. 15B, bottom track) and ~5% of all cytoplasmic sisRNAs have high

probability snoRNA motifs, confirming that slb-snoRNAs may accumulate in the cytoplasm. Importantly, we are sure of the cytoplasmic localization because the *intact* GV was manually removed before RNA purification.

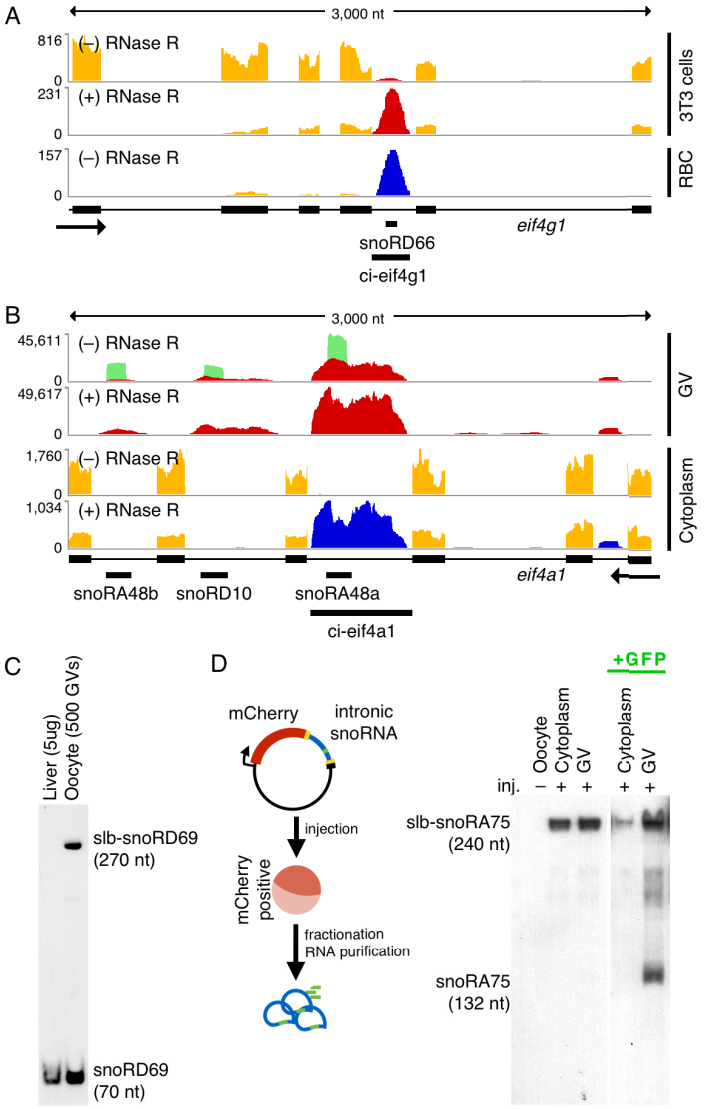


Figure 15. Some slb-snoRNAs accumulate in the cytoplasm. (A) mouse slb-snoRD66 is detected in 3T3 cells (red) and RBCs (blue). (B) Mature *Xenopus tropicalis* snoRA48 is detected in the GV sample (green) and slb-snoRA48 is detected in the GV and cytoplasm of the frog oocyte (red and blue). (C) *Xenopus tropicalis* snoRD69 (lower band) and slb-snoRD69 (higher band) are detected by northern blot. (D) *Xenopus laevis* oocytes were injected with 100pg of *Xenopus tropicalis* snoRA75 construct. Ectopic slb-snoRA75 (higher band) is detected in the GV and cytoplasm of *Xenopus laevis* oocytes by northern blot. With co-injection of 750 pg of GFP construct, slb-snoRA75 is mainly detected in the GV and snoRA75 is now readily detected in the GV.

slb-snoRNAs are exported to the cytoplasm

To our advantage, slb-snoRNA can be overexpressed in *Xenopus* oocytes. In a previous study, I had designed a DNA construct optimized to express sisRNAs ectopically (Chapter 2). Here, I injected the *Xenopus tropicalis* slb-snoRA75 sequence inserted in mCherry 3' UTR into *Xenopus laevis* oocytes (Fig. 15D). Two days after injection, both *xt*SNORA75 and slb-snoRA75 were detected in the GV but only slb-snoRA75 accumulated in the cytoplasm.

Next, I performed a competition assay to determine the export machinery of slb-snoRNAs. Consistent with our previous study (Chapter 2), slb-snoRNA cytoplasmic accumulation was diminished when oocytes were co-injected with a substantial amount of competitor plasmid encoding GFP. These results suggest that slb-snoRNA and GFP mRNA share the same export machinery, demonstrating that slb-snoRNAs do not passively leak through the nuclear envelope, but are actively exported to the cytoplasm.

It is worth noting that mature snoRA75 was more abundant relative to nuclear slb-snoRA75 in oocytes co-injected with large amount of the GFP construct (Fig 15D). These data suggest that export of slb-snoRNAs to the cytoplasm and snoRNA maturation in the nucleus are competing reactions.

slb-snoRNAs are bound to canonical snoRNP proteins

Next, I asked if slb-snoRNAs were associated with snoRNP proteins. The RNA binding protein of H/ACA box snoRNAs is dyskerin, the catalytic protein of the H/ACA snoRNP. To test if dyskerin associates with sln-snoRNAs, I expressed slb-snoRA75 alone or together with HA-tagged dyskerin in *Xenopus laevis* oocytes and carried out immunoprecipitation experiments (Fig. 16A). In both samples, slb-snoRA75 was detected

by northern blotting in input as expected. In the pull-down samples however, slb-snoRA75 was detected only when precipitating dyskerin-HA and not in the control (Fig. 16B), suggesting that snoRNP proteins associate with slb-snoRNAs. Next, both nuclear and cytoplasmic dyskerin-HA pool were precipitated and the RNA co-precipitated was analyzed by RT-PCR using outward facing oligos that detect circular molecules only. I found that both nuclear and cytoplasmic slb-snoRNAs are associated with dyskerin (Fig. 16C). Moreover, endogenous nuclear and cytoplasmic slb-snoRNA75 co-precipitated HA-tagged dyskerin as well (Fig. 16D), suggesting that slb-snoRNA and snoRNP proteins form a dynamic complex.

Next, I analyzed C/D box snoRNAs and particularly the C/D box RNA binding protein 15.5K which is required for the formation of a stable snoRNP complex. To test for endogenous complexes, I carried out an immunoprecipitation assay against 15.5K. Although I could pull down only a very small fraction of the endogenous protein, I could detect endogenous *Xenopus tropicalis* slb-snoRD41 that co-precipitated with 15.5K (data not shown). Altogether, the data show that slb-snoRNAs can form a stable complex with snoRNA binding proteins. Next, in collaboration with Svetlana Deryusheva, we asked whether slb-snoRNAs could carry out a canonical snoRNP function, namely, if they have guide RNA activity.

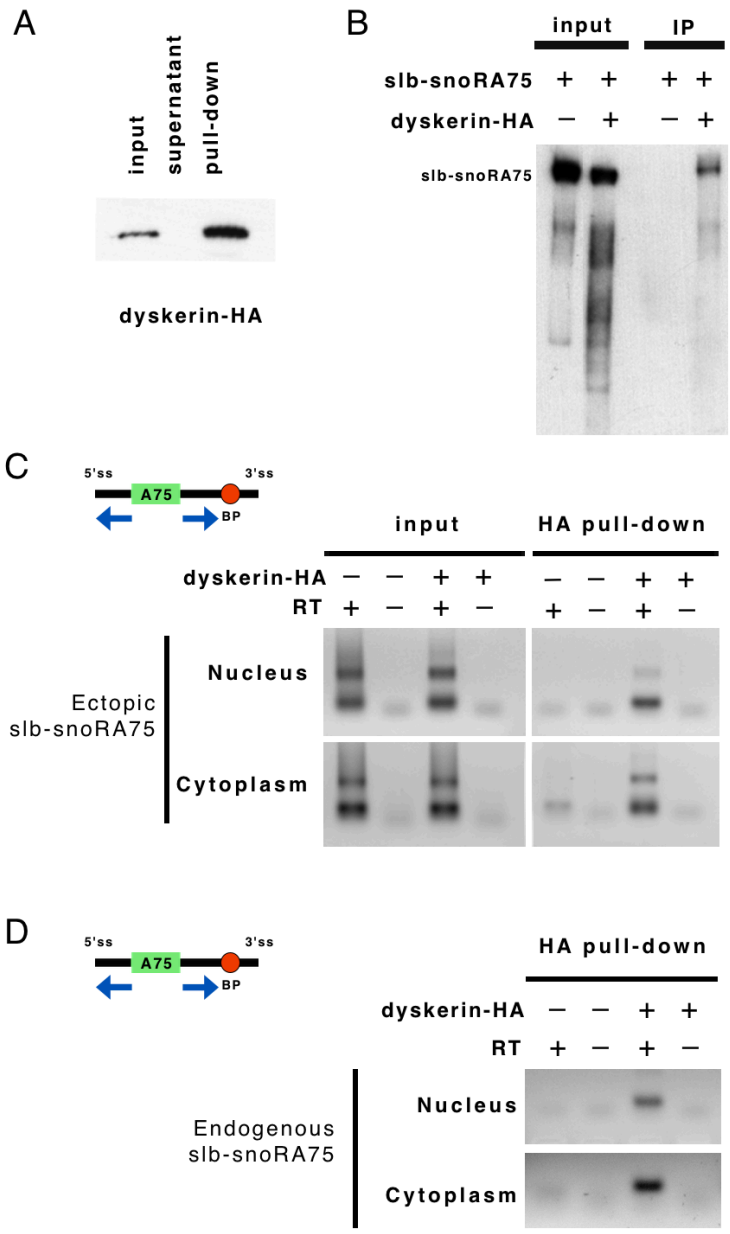


Figure 16. slb-snoRNA and snoRNP proteins form a dynamic complex. (A) *Xenopus laevis* oocytes were injected with a HA tagged dyskerin DNA construct and proteins were precipitated with HA-beads. Western blot analysis of the ectopic protein using an antibody against HA tag. (B) Northern blot analysis of slb-snoRA75 in the input and pull-down samples analyzed by western blot. Slb-snoRA75 co-precipitates with HA tagged dyskerin. (C) RT-PCR analysis using outward facing oligos against *X. tropicalis* sln-snoRA75 (ectopic). Both nuclear and cytoplasmic slb-snoRA75 co-precipitated with dyskerin. (D) RT-PCR analysis of endogenous slb-snoRA75 precipitated with ectopic dyskerin.

rRNA is modified in* dbr1Δ *and* dbr1Δrnt1Δ *mutants

To test if snoRNAs in lariat form could guide 2'-O-methylation and pseudouridylation, we switched to the yeast *S. cerevisiae* cell system. Of particular interest, the debranching enzyme mutant strain, *dbr1Δ*, is viable and accumulates lariats in the cytoplasm. In the *dbr1Δ* mutant strain, it has been shown that the guide RNA snR24 cannot be processed, but instead accumulates in the form of a lariat (Ooi *et al*., 1998). At the same time, positions targeted for modification by snR24 (C1437, C1449, and C1450 of the large subunit (LSU) of rRNA) remain modified in *dbr1*Δ. The authors concluded that snoRNAs could be functional when in a lariat form. However, along with the lariat form, partially processed snR24 transcripts were also readily detectable by northern blot analysis (Ooi *et al*., 1998). This partial processing could be driven by RNT1, an exonuclease known to be involved in snoRNA processing (Chanfreau et al. 1998). To distinguish between a functional lariat or a functional partially processed snoRNA, we analyzed the rRNA modification pattern in *dbr1Δ* and the double mutant *dbr1Δrnt1Δ*.

Dr. Svetlana Deryusheva found similar patterns with all eight intronic snoRNAs known in yeast (Fig. 17). In *dbr1Δ* and the double mutant *dbr1Δrnt1Δ*, all snoRNAs accumulated in the lariat, partially processed, and mature forms, although the latter two are at low levels (Fig. 17A). Consistent with the observations of Ooi *et al*. (1998), the targeted positions were modified in wild-type and both mutant strains (Fig. 17B). Notably, independently transcribed snoRNAs were barely detectable in *dbr1Δrnt1Δ*, yet we could detect the targeted modifications (Fig. 17B and C). Because there is no alternative mechanism to modify these positions, we concluded that a minimal amount of

mature snoRNA is sufficient for modification. However, we cannot rule out that the detected rRNA modifications were guided by some mature snoRNA processed from the intronic RNA in *dbr1Δ* and *dbr1Δrnt1Δ*.

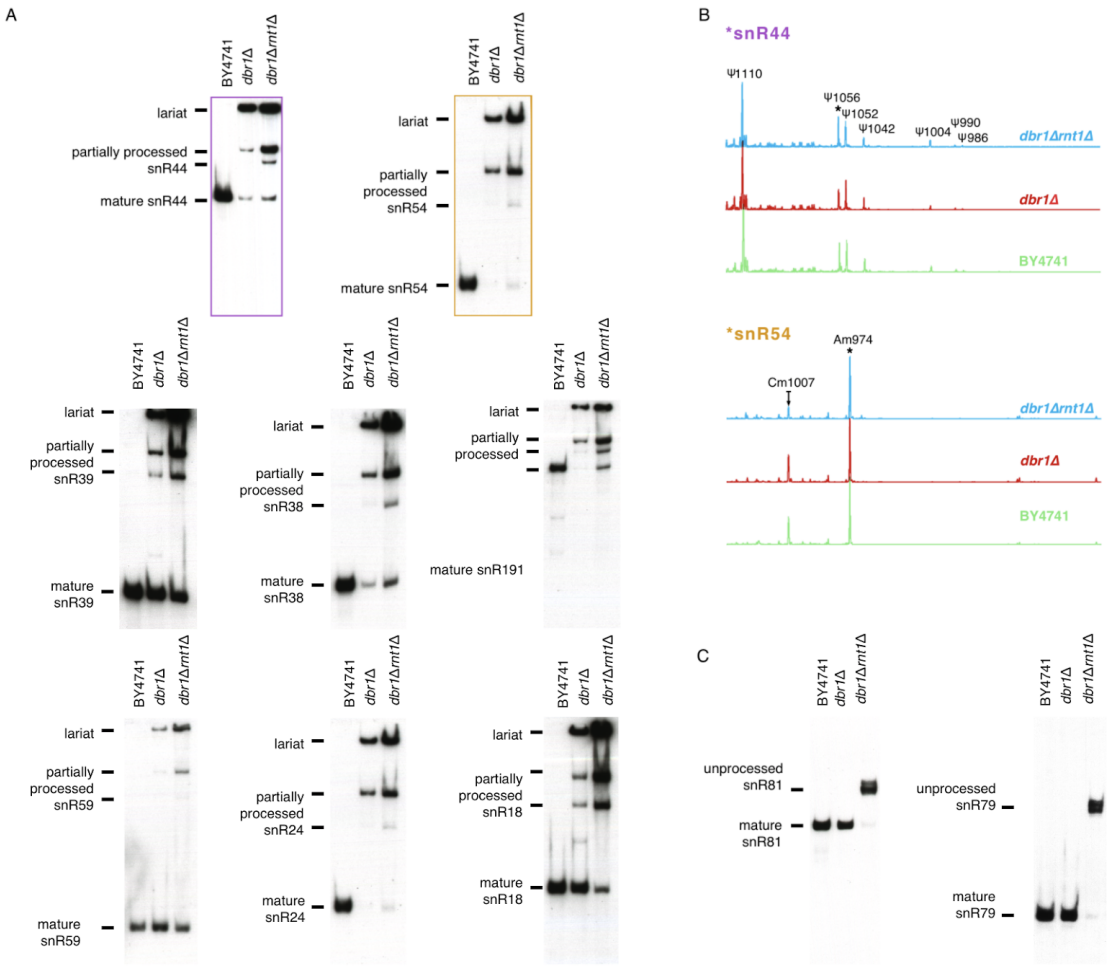


Figure 17. Endogenous guide RNAs in *dbr1Δ* and *dbr1Δrnt1Δ*. (A) Northern blot analysis of all intronic snoRNAs carried out by Dr. Svetlana Deryusheva. (B) rRNA modification pattern detected in yeast *dbr1Δ* and *dbr1Δrnt1Δ* by Dr. Svetlana Deryusheva. (C) Northern blot analysis of snoRNA encoded in independent transcripts in both mutant. In the case of H/ACA box snoRNAs, LSU-U1054 (targeted by snR44) was modified in both mutants. A slight decrease in pseudouridylation of U1054 in the *dbr1Δ* mutant correlates with decrease of mature snR44. Both the non-intronic guide, snR81, and its target, LSU-U1052, are reduced in *dbr1Δrnt1Δ*. In the case of C/D box snoRNAs, the non-intronic guide snR79 and its target are also decreased in *dbr1Δrnt1Δ*. However, the mature form of the intronic guide snR54 was

decreased but not the methylation it targets. Intronic C/D box snoRNAs could be functional in the lariat form or as a partially processed transcript to compensate for the decrease of the mature form, as previously observed.

slb-snoRNAs do not function as post-transcriptional modification guides in yeast

To eliminate the accumulation of mature and partially processed intronic snoRNAs, we expressed *Xenopus tropicalis* slb-snoRA28 and slb-snoRD41 in yeast. We inserted the *Xenopus* slb-snoRNAs intron in the yeast EFB1 gene, in place of the intron encoding snR18, and the essential splicing motifs were replaced with yeast canonical sequences. As a control, we expressed snoRA28 and snoRD41 as independent genes (Fig. 18A). Together, these conditions were sufficient to abolish mature intronic snoRNA accumulation in *dbr1Δ* or *dbr1Δrnt1Δ*.

When snoRA28 was expressed as an independent transcript in *dbr1Δ,* mature snoRA28 was barely detected by northern blot, yet SSU-U808 was modified in these cells (Fig. 18B, green curve). On the other hand, when intronic snoRA28 was expressed, processing of the guide was impaired in *dbr1Δ* and SSU-U808 was not modified (Fig. 18B, red curve). Instead of the mature form, we detected circular slb-snoRA28 and linearized slb-snoRA28 in the *dbr1Δ* mutant. Even though the lariat and linear 'partially processed' transcripts were abundant, they were not sufficient to modify SSU-U808. Altogether, these data suggest that box H/ACA snoRNAs are not functional in a lariat form.

To analyze box C/D snoRNA, we expressed *Xenopus tropicalis* slb-snoRD41. Because the canonical target of snoRD41, LSU-U2729, is modified by endogenous machinery, we altered the antisense element (ASE) so that the guide targets U2 snRNA at position C41 instead. In *dbr1Δ* transformed with the chimeric snoRD41-independent

transcript construct, we detected low levels of the mature chimeric guide and modification at U2-C41 (Fig. 18C, dark green curve). Strikingly, we could detect the modification even when the level of the chimeric snoRNA was extremely low (only detectable by overexposing the blot) (Fig. 18C, light green curve). Next, we analyzed the intronic chimeric guide. In the transformed *dbr1Δ* mutant, U2-C41 was modified, whereas the mature snoRNA was not evident. However, when expressed in the *dbr1Δrnt1Δ* double mutant, the intronic chimeric snoRNA accumulated *only* as a lariat and U2-C41 was not modified. Because the levels of lariats in the *dbr1Δrnt1Δ* strain were 3 to 5 times higher than the levels of 'low-level' chimeric guide expressed as an independent transcript, the absence of modification can be explained only by the absence of snoRNA guide activity in the lariat form. Based on all these experiments, we conclude that slb-snoRNAs cannot guide post-transcriptional modifications.

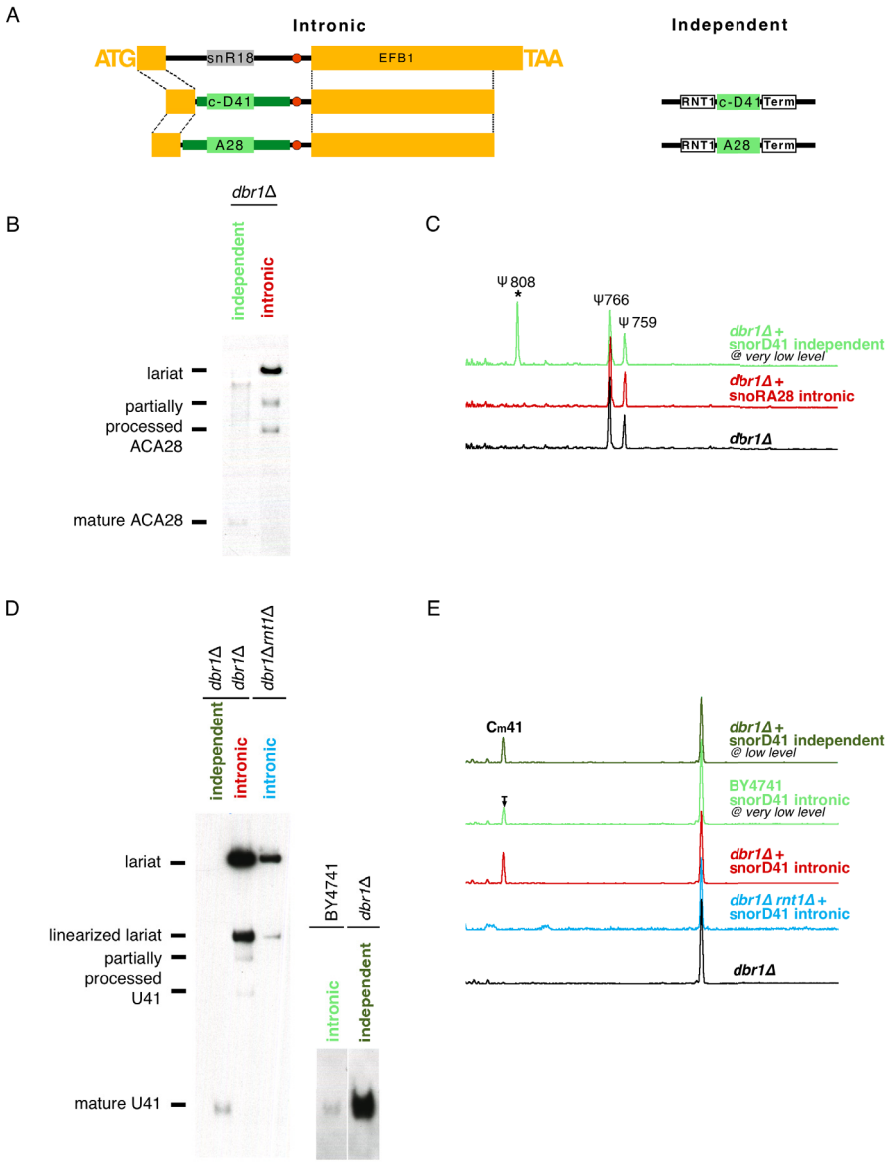


Figure 18. slb-snoRNAs do not modify their canonical target in yeast. (A) Schematic of the DNA constructs transformed in yeast *dbr1Δ* and *dbr1Δrnt1Δ*. (B) In *dbr1Δ,* mature snoRA28 (ACA28) is detected by northern blot only when the guide is expressed from an independent transcript (Dr. Svetlana Deryusheva). (C) Partial small subunit (SSU) rRNA modification mapping done by Svetlana Deryusheva. Ectopic expression of snoRA28 from an independent transcript was sufficient to modify SSU-U808 in *dbr1Δ*. In the same mutant, intronic snoRA28 cannot modify SSU-U808. (D) Northern blot analysis carried out by Svetlana Deryusheva and Gaëlle Talhouarne against chimeric snoRD41 (U41). In *dbr1Δ,* the independent chimeric snoRD41 gene produces the mature form of the guide, while the intronic snoRD41 construct leads to accumulation of a partially processed lariat and a lariat. In *dbr1Δrnt1Δ* the intronic chimeric snoRD41 construct generates lariats only. (E) U2 modification mapping done by Svetlana Deryusheva. Very low levels of mature snoRD41 are sufficient to modify U2-C41; partially processed lariats could modify U2-C41, but lariats could not.

DISCUSSION

slb-snoRNA vs. circular snoRNA

In this chapter, we establish that snoRNAs can accumulate as lariats in human, mouse, chicken, and frog cell cultures. We observed that nearly all snoRNAs accumulate as mature guides and slb-snoRNAs in the GV of the *Xenopus* oocytes. Furthermore, a subset of slb-snoRNAs can be exported to the cytoplasm. Based on ectopic expression of *Xenopus* slb-snoRNAs in yeast, we conclude that slb-snoRNAs cannot function as canonical modification guide RNAs.

In Archaea, however, guide RNAs may be encoded in tRNA introns (although rare) and, once spliced by an endonuclease, the RNA is circularized by a splicing ligase. Such a circular guide RNA is functional (Salgia *et al*., 2003; Starostina *et al*., 2004; Danan M *et al*., 2012).

The difference in functionality may be due to key differences between the two circular transcripts. First, eukaryotic slb-snoRNAs are at least ~100 nt longer than mature snoRNAs, whereas archaeal circular guides are only 2 or 3nt longer. It is possible that the bulkiness of the slb-snoRNA prevents its function as a guide. Moreover, eukaryotic and archaeal circular introns have different junctions. In lariats, the branchpoint is linked to the 5'ss by a 5'-2' covalent bond whereas archaeal tRNA introns are ligated at a canonical 5'-3'. The 5'-2' link may accentuate the bulkiness of slb-snoRNAs. Finally, in eukaryotes, rRNA/snRNA post-transcriptional modifications occur in the nucleus where snoRNAs accumulate, whereas Archaea do not have nuclei. Of significance, in *dbr1Δ*, lariats have been shown to accumulate in the cytoplasm, presumably to reduce the toxicity of the unprocessed transcripts in the nucleus. Therefore, it is possible that slb-

snoRNAs cannot modify rRNA and snRNA because they are rapidly exported to the cytoplasm in yeast, whereas Archaea would not have this barrier.

These differences may explain why slb-snoRNAs cannot modify their targets. However, we cannot rule out that slb-snoRNA may modify non-canonical targets that are more 'open' and in the same compartment as slb-snoRNAs.

Export of slb-snoRNA and implications

I have determined that sisRNAs are exported by NXF1/NTF1 (Chapter 2). In this study, I find that slb-snoRNAs are exported by NXF1/NXT1 as well. Additionally, I found that slb-snoRNA export prevents snoRNA maturation. Because the debranching enzyme is strictly nuclear in frog oocytes (Chapter 1), export to the cytoplasm would be the most efficient way to prevent linearization and further processing, consequently preventing snoRNA maturation.

On the other hand, mature snoRNAs can accumulate in the cytoplasm of cultured rat cells, but only under stress conditions (Holley *et al*., 2015). The export mechanism remains unknown. One intriguing possibility is that slb-snoRNAs may be exported to the cytoplasm and, under stress conditions, the lariats may be further processed by a non-canonical pathway to initiate snoRNA maturation.

slb-snoRNA may add a layer of regulation for post-transcriptional modification

Normal cell homeostasis requires tight control of snoRNA levels; in fact, many diseased cells show deregulation of snoRNP RNA and proteins (Chang *et al*., 2002; Dong *et al*., 2008, 2009; Lia *et al*.,2010; Mannoor *et al*., 2014; Flockhart *et al*.,2016; Yoshida *et al*., 2017). Therefore, cells must have mechanisms to buffer changes in snoRNA levels

that may occur during stress or diseased conditions. It is possible that slb-snoRNAs are formed to prevent excessive accumulation of mature snoRNAs. Alternatively, slb-snoRNAs may allow for rapid snoRNA production. That is to say, the stable lariat bound to the snoRNP proteins requires only linearization and further processing to make a functional snoRNA. It remains unclear whether the cytoplasmic slb-snoRNAs can be imported back to the nucleus as intact circular molecules or, once processed, as functional snoRNAs.

Because many stable lariats have snoRNA-like motifs rather than functional snoRNAs (Fig. S3), they may be 'sponging' excessive snoRNP proteins. A similar phenomenon has been described in plants. Li *et al*. (2016) recently found that stable lariats can deplete the pool of dicer, which results in a general down-regulation of miRNAs. Overall, stabilization of lariats that bear a snoRNA motif adds a new layer of regulation for cells to fine-tune post-transcriptional processing of other RNAs.

MATERIALS AND METHODS

RNA purification

RNA was extracted with TRIzol (Ambion) followed by Direct-zol RNA MiniPrep kit (Zymo Research). DNA was removed on the columns according to the manufacturer's protocol. The RNA was quantified with a Nanodrop One (Thermo Scientific).

RT-PCR

Reverse transcription was carried using random hexamers with Episcript (Epicentre) and AMV-RT (New England Biolabs) for 1 hr at 37℃. cDNA was amplified with inward and outward facing primers (Table S3, Appendix) using Taq polymerase. To confirm that sisRNAs were circular, the amplified DNAs were cloned into the pGEM-T Easy vector (Promega) and analyzed by Sanger sequencing.

Animals, oocytes and microinjections

Xenopus laevis females were anesthetized with tricaine methanesulfonate (MS222), pH 7.0, and a portion of their ovary was removed. After manual separation, the largest cells were cultured in OR2 medium at 16℃. Injection methods were previously described in Chapter 2. In short, oocytes were injected at the animal pole with 100 pg of plasmid in a 2.4nl volume. 48 hrs after injection, nuclei were dissected from the oocytes in a pH 5.6 isolation solution. Nuclear and cytoplasmic RNAs were extracted with TRIzol.

Immunoprecipitation

Oocytes were dissected in OR2 isolation buffer at pH 7.0 (83 mM KCl, 17 mM NaCl, 6.0 mM Na2HPO4, 4.0 mM KH2PO4, 1 mM MgCl2, 1.0 mM dithiothreitol (DTT). Cytoplasms and GVs were collected in separate tubes and homogenized in lysis buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 5 mM MgCl2, 1 mM DTT with a proteinase inhibitor cocktail (Roche)). The lysates were homogenized by drawing up and down in a 27-gauge needle and cleared by centrifugation for 10 min at 14 krpm. The supernatants were collected and incubated with HA beads for 1 hr at room temperature. Alternatively, cleared lysates were incubated with antibodies against 15.5K (kindly provided by Dr. Lührmann) for 1 hr and an additional 1 hr with A/G magnetic beads (Biotools). The proteins were eluted in 1X reducing sample buffer and RNA was extracted with TRIzol. Samples were boiled and thoroughly shaken before the beads were removed.

Northern blot

RNA was separated on an 8% polyacrylamide gel, transferred onto a nylon membrane (Zeta Probe, Bio-Rad), and probed with digoxigenin-labeled PCR probes. Hybridization was carried at 40°C in DIG Easy Hyb buffer (Roche) and probes were detected with digoxigenin antibody conjugated with AP and the chemiluminescent substrate CDP-Star (Roche).

Western blot

Proteins were separated on a 12.5% acrylamide gel and transferred onto a PVDF membrane (Millipore). Proteins were detected with rat anti-HA antibody 3F10 (Roche) (40 ng/ml) and rabbit anti-15.5K (1:5,000) for 1 hr at RT followed by HRP-conjugated

secondary antibodies (diluted in 5% milk at 1:20,000). HRP was detected with SuperSignal West Dura chemiluminescent substrate (Thermo Scientific).

Yeast strains and media

Haploid yeast *S. cerevisiae* strains used in this study were the following: BY4741 (wild type control), the *dbr1*Δ*, DBR1::KAN* mutant strain (kindly provided by Jeff Han), and the double knockout *dbr1*Δ*rnt1*Δ*, DBR1::HIS RNT1::TRP* strain (kindly provided by Sherif Abou Elela, Université de Sherbrooke, Canada).

To express exogenous RNAs in yeast cells, we made constructs from yeast expression vectors p426Gal1, p416GalS (kindly provided by Jeff Han) and YEplac195 [URA3 2m] containing a *GPD* promoter, an RNT1 cleavage site, and an *snR13* terminator generated in the Yi-Tao Yu Lab (Huang *et al*., 2011). Inserted fragments were amplified by PCR from yeast and *Xenopus* genomic DNA. Corresponding restriction sites were added with primers. Overlap-extension PCR was used to make chimeric fragments. Schematic representation of these constructs is shown in Fig.17A.

Yeast cells were transformed using standard lithium acetate methods. Yeast strains were grown in rich YPD medium; transformants were grown in synthetic selective media with either glucose or galactose (in the case of an inducible *Gal* promoter) as a source of sugar at 30° C or at 25° C (when the *dbr1*Δ*rnt1*Δ mutant strain was used in the experimental setup).

Fluorescent Primer Extension Analysis to Map 2'-O-Methylation and Pseudouridylation

For mapping 2 -*O*-methylation and pseudouridylation sites in yeast U2 snRNA and rRNAs, we used a nonradioactive modification of methods based on reverse transcription (Deryusheva *et al.*, 2012). Oligonucleotides terminally labeled with a 6-FAM fluorescent dye were used as primers; 3–6 pmol of 6-FAM-oligo was used per reaction. Modification mapping reactions were performed in a 20 ml volume using AMV-RT (New England Biolabs) at 42°C.

To detect 2'-*O*-methylation, we performed primer extension with a low concentration of deoxynucleotide triphosphate (dNTP); the dNTP concentration varied from 0.004 to 0.01 mM and was optimized for each RNA. Control reactions were done at a high concentration of dNTPs (0.5 mM).

To detect pseudouridines, test RNA samples were treated first with CMC [N-cyclohexyl-N′ -(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate] (Sigma-Aldrich) for 20–30 min and then with pH 10.4 sodium carbonate buffer for 3–4 h at 37°C. Primer extension was performed at 0.5 mM dNTP concentration. Untreated RNA was used as a control.

RNA sequencing products were used to determine the exact position of modified nucleotides. Chain termination mixes contained 270 μM dNTPs, where one dNTP was partially replaced with acyNTP (New England Biolabs, Beverly, MA) at a 1:3 (dNTP:acyNTP) ratio.

Reaction products were purified by ethanol precipitation, lyophilized, and kept at -20°C until use. Dry pellets were dissolved in formamide, mixed with the GeneScan-500

LIZ Size Standard and separated on an ABI3730xl capillary electrophoresis instrument (Applied Biosystems) using techniques and parameters suggested by the manufacturer. The Gene Scan-500 Liz Size Standard was included in each run to align fragments from different samples.

GeneMapper version 3.7 software (Applied Biosystems) was used to screen the data, identify peaks, and precisely determine positions of modified nucleotides. PeakScanner, a free software, was also used for this analysis.

REFERENCES

Chanfreau G, Rotondo G, Legrain P and Jacquier A. **1998**. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *The EMBO Journal*, *17*(13), 3726–3737.

Chang, L.S., Lin, S.Y., Lieu, A.S., Wu, T.L., **2002**. Differential expression of human 5S snoRNA genes. Biochem. Biophys. Res. Commun. 299, 196–200.

Danan M, Schwartz S, Edelheit S, and Sorek R. **2012**. Transcriptome-wide discovery of circular RNAs in Archaea. Nucleic Acids Res 40: 3131–3142.

Deryusheva S, Choleza M, Barbarossa A, Gall JG, Bordonné R. **2012**. Post-transcriptional modification of spliceosomal RNAs is normal in SMN-deficient cells. RNA. 18(1):31-6.

De Araujo Oliveira, J. V., Costa, F., Backofen, R., Stadler, P. F., Machado Telles Walter, M. E., and Hertel, J. **2016**. SnoReport 2.0: new features and a refined Support Vector Machine to improve snoRNA identification. *BMC Bioinformatics*, *17*, 464.

Dong, X.Y., Rodriguez, C., Guo, P., Sun, X., Talbot, J.T., Zhou, W., Petros, J., Li, Q., Vessella, R.L., Kibel, A.S., *et al*., **2008**. SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. Hum. Mol. Genet. 17, 1031–1042.

Dong, X.Y., Guo, P., Boyd, J., Sun, X., Li, Q., Zhou, W., Dong, J.T. **2009**. Implication of snoRNA U50 in human breast cancer. J. Genet. Genomics 36, 447–454.

Flockhart, R., Zarnegar, B.J., Che, Y., Meschi, F., *et al*., **2016**. The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. Nat. Genet. 48, 53–58.

Holley CL, Li MW, Scruggs BS, Matkovich SJ, Daniel S. Ory Schaffer JE. **2015**. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell Metab*;14:33–44

Li Z, Wang S, Cheng J, Su C, Zhong S, Liu Q, Yuda Fang Y, Yu Y, Lv H, Zheng Y and Zheng B. **2016**. Intron Lariat RNA Inhibits MicroRNA Biogenesis by Sequestering the Dicing Complex in *Arabidopsis.* PLoS Genet. 12(11): e1006422.

Liao, J., Yu, L., Mei, Y., Guarnera, M., Shen, J., Li, R., Liu, Z., Jiang, F. **2010**. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Mol. Cancer 9, 198.

Mannoor, K., Shen, J., Liao, J., Liu, Z., Jiang, F. **2014**. Small nucleolar RNA signatures of lung tumor-initiating cells. Mol. Cancer 13, 104.

Ooi, S.L., Samarsky, D.A., Fournier, M.J., and Boeke, J.D. **1998**. Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. RNA 4, 1096–1110.

Petfalski E, Dandekar T, Henry Y and Tollervey D. **1998**. Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol Cell Biol*, 18,1181–1189.

Starostina NG, Marshburn S, Johnson LS, Eddy SR, Terns RM, and Terns MP. **2004**. Circular box C/D RNAs in *Pyrococcus furiosus*. Proc Natl Acad Sci 101: 14097–14101.

Salgia SR, Singh SK, Gurha P, and Gupta R. **2003**. Two reactions of *Haloferax volcanii* RNA splicing enzymes: joining of exons and circularization of introns. RNA 9: 319–330.

Yoshida K, Toden S, Weng W, Shigeyasu K, Miyoshi J, Turner J, Nagasaka T, Ma Y, Takayama T, Fujiwara T, Goel A. **2017**. SNORA21 - An Oncogenic Small Nucleolar RNA, with a Prognostic Biomarker Potential in Human Colorectal Cancer. EBioMedicine. pii: S2352-3964(17)30282-7.

Yu, Y.-T., Terns, R.M., Terns, M.P. **2005**. *Fine-tuning of RNA modifications by modification and editing*, Mechanisms and functions of RNA-guided RNA modification, ed Grosjean H.(Springer, New York), pp 223–262.

Table S1. List of sisRNAs with snoRNA motifs

genome	sisRNA locus	snoRNA motif	probability	annotation	genome	sisRNA locus	snoRNA motif	probability	annotation
hg19	chr1:45241812-45242346(+)	C/D box	1.00	**snoRD46**	Xenla9	Scaffold54:479747-479937(-)	C/D box	1.00	**snoRD56**
hg19	chr19:51305514-51305710(-)	C/D box	1.00	**snoRD88**	Xenla9	Scaffold54:479111-479625(-)	C/D box	1.00	**snoRD57**
hg19	chr19:49994121-49994296(+)	C/D box	0.99	**snoRD34**	Xenla9	chr5L:19434097-19434601(-)	C/D box	1.00	**snoRD92**
hg19	chr16:30678311-30678640(+)	H/ACA box	0.967412		Xenla9	chr4S:22260169-22260652(+)	C/D box	1.00	**snoRD14**
hg19	chr22:40805022-40805253(+)	H/ACA box	0.957347		Xenla9	chr6L:2946949-2947317(+)	H/ACA box	0.999968	**snoRA16**
hg19	chr19:10220383-10220588(+)	C/D box	0.95	**snoRD105**	Xenla9	chr7L:110109839-110110078(+)	H/ACA box	0.993513	
hg19	chr3:52524247-52524720(+)	C/D box	0.94		Xenla9	chr8L:390497-390967(-)	H/ACA box	0.992932	
hg19	chr16:57092973-57093378(+)	C/D box	0.92		Xenla9	chr3L:133157285-133157474(+)	C/D box	0.99	
hg19	chr2:27668316-27668611(-)	H/ACA box	0.904722		Xenla9	chr1S:483518-483739(+)	C/D box	0.99	**snoRD110**
hg19	chr1:26610973-26611954(-)	C/D box	0.88		Xenla9	chr9_10L:8459871-8460171(-)	C/D box	0.99	**snoRD12**
hg19	chr11:71942233-71942541(+)	H/ACA box	0.811045		Xenla9	chr9_10L:8459185-8459454(-)	C/D box	0.99	**snoRD12**
hg19	chr3:52028142-52029057(-)	H/ACA box	0.788859		Xenla9	chr9_10S:7563563-7564029(-)	C/D box	0.99	**snoRD12**
hg19	chr17:43012303-43012626(-)	H/ACA box	0.753392		Xenla9	chr2L:87841018-87841333(+)	C/D box	0.99	
hg19	chr11:64081839-64082212(+)	H/ACA box	0.721933		Xenla9	chr8S:32287294-32287824(-)	C/D box	0.99	**snoRD62**
hg19	chr2:27456008-27456179(+)	H/ACA box	0.638971		Xenla9	chr9_10L:8459516-8459807(-)	C/D box	0.98	**snoRD12**
hg19	chr17:17079826-17080563(+)	C/D box	0.55		Xenla9	chr2L:76921057-76921504(+)	H/ACA box	0.977109	
hg19	chr8:145193716-145193911(+)	H/ACA box	0.500000		Xenla9	chr5L:140914269-140914730(+)	H/ACA box	0.976991	
hg19	chr8:145438924-145439119(+)	H/ACA box	0.500000		Xenla9	Scaffold36:1956778-1957240(-)	H/ACA box	0.976645	
mm10	chr2:130277970-130278155(+)	C/D box	0.99	**snoRD57**	Xenla9	chr9_10L:44602783-44602972(-)	C/D box	0.97	**snoRD33**
mm10	chr2:167065217-167065433(+)	C/D box	0.99	**snoRD12**	Xenla9	chr5L:19431347-19431699(-)	C/D box	0.97	**snoRD53**
mm10	chr7:45126794-45126990(-)	C/D box	0.99	**snoRD33**	Xenla9	chr8L:16180582-16180804(-)	C/D box	0.97	**snoRD117**
mm10	chr19:6253068-6253333(+)	H/ACA box	0.983447		Xenla9	chr2L:173884682-173885117(-)	H/ACA box	0.969321	
mm10	chr15:102710530-102710913(-)	H/ACA box	0.970961		Xenla9	chr9_10L:44942644-44943316(+)	H/ACA box	0.967168	
mm10	chr6:84107002-84108046(+)	C/D box	0.97		Xenla9	chr8S:26297229-26297507(-)	C/D box	0.96	
mm10	chr2:26912445-26912670(+)	C/D box	0.94	**snoRD36**	Xenla9	chr9_10S:94922002-94922223(+)	H/ACA box	0.958628	
mm10	chr19:5759342-5759664(-)	H/ACA box	0.935440		Xenla9	chr7L:4911753-4912071(-)	H/ACA box	0.951726	
mm10	chr11:78167886-78168091(-)	H/ACA box	0.887478		Xenla9	chr9_10L:114860279-114860451(-)	H/ACA box	0.951321	
mm10	chr2:125257140-125257356(+)	C/D box	0.88		Xenla9	Scaffold54:480541-480725(-)	C/D box	0.95	**snoRD110**
mm10	chr1:73941053-73941909(-)	C/D box	0.87		Xenla9	Scaffold20:2210138-2210335(-)	C/D box	0.95	
mm10	chrX:7729933-7730267(+)	H/ACA box	0.844504		Xenla9	chr8L:7389575-7390036(+)	H/ACA box	0.949798	
mm10	chr4:62304906-62305122(-)	H/ACA box	0.835113		Xenla9	chr7S:4284712-4284972(-)	H/ACA box	0.946281	
mm10	chr8:95097764-95097954(+)	H/ACA box	0.827305		Xenla9	chr2S:28781364-28781656(-)	H/ACA box	0.945955	
mm10	chr6:37335954-37336214(-)	H/ACA box	0.800142		Xenla9	chr5L:23376838-23377205(+)	H/ACA box	0.942807	
mm10	chr6:115965971-115966493(-)	H/ACA box	0.791814		Xenla9	chr1L:217504163-217504348(+)	H/ACA box	0.941349	
mm10	chr17:35099000-35100334(+)	C/D box	0.79		Xenla9	Scaffold164:37587-38088(-)	H/ACA box	0.940137	
mm10	chr7:116333098-116333695(-)	H/ACA box	0.782991	**snoRD14**	Xenla9	chr9_10L:113741061-113741424(-)	H/ACA box	0.936952	
mm10	chr15:81705748-81706400(-)	H/ACA box	0.776038		Xenla9	chr6L:2948477-2948795(+)	C/D box	0.93	**snoRD99**
mm10	chr9:26977658-26978310(-)	H/ACA box	0.746350		Xenla9	chr7L:121207866-121208091(+)	H/ACA box	0.928052	
mm10	chr16:17095043-17096004(-)	H/ACA box	0.741503		Xenla9	chr9_10S:38476748-38477024(+)	H/ACA box	0.925222	
mm10	chr7:19581158-19581380(-)	H/ACA box	0.734778		Xenla9	chr6L:151799709-151799921(-)	H/ACA box	0.915099	
mm10	chr2:180033870-180034297(+)	H/ACA box	0.714631		Xenla9	chr1L:7382721-7383133(+)	H/ACA box	0.910434	
mm10	chr7:28814287-28815098(+)	H/ACA box	0.695794		Xenla9	Scaffold20:4892314-4892530(-)	H/ACA box	0.905335	
mm10	chrX:101542682-101543050(+)	C/D box	0.67		Xenla9	chr3S:119435352-119435661(+)	C/D box	0.89	
mm10	chr11:69089226-69089730(+)	C/D box	0.65		Xenla9	chr3S:18881581-18881917(-)	H/ACA box	0.889047	
mm10	chr9:108948633-108948938(+)	H/ACA box	0.635735		Xenla9	chr2L:1383915-1384150(+)	H/ACA box	0.877655	
mm10	chr4:155731464-155731926(+)	C/D box	0.63		Xenla9	chr9_10L:33335417-33335907(-)	H/ACA box	0.876179	
mm10	chr11:102915700-102916188(-)	H/ACA box	0.606623		Xenla9	chr2L:172231344-172231869(-)	H/ACA box	0.860142	
mm10	chr15:78903135-78903552(+)	H/ACA box	0.585385		Xenla9	chr8L:29423521-29424243(-)	H/ACA box	0.834723	
mm10	chr11:120822230-120822716(-)	C/D box	0.58		Xenla9	chr3S:4512407-4512693(+)	H/ACA box	0.830322	
mm10	chr1:59167279-59167462(-)	H/ACA box	0.578924		Xenla9	chr8L:605741-606067(+)	H/ACA box	0.826425	
mm10	chr14:121276505-121276898(+)	H/ACA box	0.572364		Xenla9	Scaffold124:145723-145986(+)	H/ACA box	0.824215	
mm10	chr8:71323033-71323264(+)	H/ACA box	0.558180		Xenla9	chr9_10S:2205838-2206205(+)	H/ACA box	0.815291	

genome	sisRNA locus	snoRNA motif	probability	annotation	genome	sisRNA locus	snoRNA motif	probability	annotation
mm10	chr7:143570149-143570541(-)	C/D box	0.50		Xenla9	chr2L:181040183-181040874(-)	H/ACA box	0.796905	
galGal5	chr24:3076518-3076829(+)	C/D box	1.00	snoRD14	Xenla9	chr8L:24734973-24735317(+)	C/D box	0.79	
galGal5	chr19:5197607-5197869(-)	H/ACA box	0.997981		Xenla9	chr9_10S:16771221-16771781(-)	H/ACA box	0.773938	
galGal5	chrZ:9182984-9183269(+)	H/ACA box	0.997784		Xenla9	chr9_10S:15121172-15121591(-)	H/ACA box	0.770300	
galGal5	chr23:5349546-5349717(+)	H/ACA box	0.996110		Xenla9	chr7S:5687840-5688043(+)	H/ACA box	0.752910	
galGal5	chr5:50298963-50299208(+)	H/ACA box	0.994713		Xenla9	chr4S:121010876-121011075(-)	H/ACA box	0.746328	
galGal5	chr11:1819969-1820612(+)	H/ACA box	0.980695		Xenla9	chr8L:116343611-116343954(+)	H/ACA box	0.740059	
galGal5	chr15:5734393-5734724(+)	C/D box	0.98		Xenla9	chr8L:119854164-119854504(-)	H/ACA box	0.739096	
galGal5	chr21:4853192-4853578(-)	C/D box	0.98		Xenla9	chr9_10S:34355319-34355922(+)	H/ACA box	0.735060	
galGal5	chr2:4380872-4381116(-)	H/ACA box	0.952864		Xenla9	chr3L:106509531-106509728(+)	H/ACA box	0.734051	
galGal5	chr25:2681191-2681759(+)	H/ACA box	0.944453		Xenla9	chr9_10L:93060476-93060968(+)	H/ACA box	0.732952	
galGal5	chr21:4911233-4911687(-)	H/ACA box	0.938089		Xenla9	chr9_10L:110346916-110347156(+)	H/ACA box	0.732295	
galGal5	chr3:96270198-96270849(+)	H/ACA box	0.930847		Xenla9	chr6S:2517010-2517326(+)	C/D box	0.73	
galGal5	chr11:18194727-18195428(-)	C/D box	0.93		Xenla9	chr8L:113357919-113358382(+)	C/D box	0.72	
galGal5	chr9:15327037-15327285(-)	C/D box	0.93	snorD82	Xenla9	chr4L:26659704-26660109(-)	H/ACA box	0.719197	
galGal5	chr11:1536992-1537355(-)	H/ACA box	0.916654		Xenla9	chr1L:205273994-205274799(+)	H/ACA box	0.708378	
galGal5	chr1:77438451-77438792(+)	H/ACA box	0.906121		Xenla9	chr8S:26804141-26804691(+)	H/ACA box	0.694168	
galGal5	chr26:1241181-1241545(+)	C/D box	0.87		Xenla9	chr4L:128699808-128700219(-)	H/ACA box	0.682821	
galGal5	chr33:417463-417944(+)	H/ACA box	0.852360		Xenla9	chr6L:23757597-23758002(-)	H/ACA box	0.682092	
galGal5	chrUn_NT_464166v1:24512-25005(+)	H/ACA box	0.847396		Xenla9	chr2S:28840637-28841449(+)	H/ACA box	0.670789	
galGal5	chr3:30994487-30995014(-)	H/ACA box	0.840560		Xenla9	chr2S:28783311-28783546(-)	H/ACA box	0.667521	
galGal5	chr27:5574954-5575350(-)	H/ACA box	0.834834		Xenla9	chr9_10L:28893373-28893845(+)	H/ACA box	0.660701	
galGal5	chr27:5085646-5085971(-)	C/D box	0.83		Xenla9	chr8L:8089861-8090252(+)	H/ACA box	0.659268	
galGal5	chr7:22151994-22152289(+)	H/ACA box	0.786645		Xenla9	chr8L:91260656-91261053(-)	H/ACA box	0.658510	
galGal5	chr25:1793532-1793973(-)	C/D box	0.76		Xenla9	chr2L:173483921-173484141(+)	H/ACA box	0.657495	
galGal5	chr5:27883026-27883352(-)	H/ACA box	0.699109		Xenla9	Scaffold36:1541606-1541809(-)	H/ACA box	0.654789	
galGal5	chr27:1567759-1568111(-)	H/ACA box	0.697053		Xenla9	chr7S:85517922-85518124(-)	H/ACA box	0.617292	
galGal5	chr5:23951146-23951640(-)	H/ACA box	0.662563		Xenla9	chr8S:28239567-28239978(+)	H/ACA box	0.616009	
galGal5	chr21:3213005-3213393(-)	H/ACA box	0.648009		Xenla9	chr9_10L:114042374-114042731(+)	H/ACA box	0.614744	
galGal5	chr28:1281710-1282336(+)	H/ACA box	0.600724		Xenla9	chr6L:154086840-154087725(+)	H/ACA box	0.613947	
galGal5	chr4:46654388-46654790(+)	H/ACA box	0.584158		Xenla9	chr9_10L:5172855-5173172(+)	C/D box	0.61	
galGal5	chr25:2839178-2839552(+)	C/D box	0.57		Xenla9	chr8L:116354304-116354822(-)	H/ACA box	0.593322	
galGal5	chr11:2194724-2195900(+)	H/ACA box	0.555423		Xenla9	Scaffold29:3128695-3129059(-)	H/ACA box	0.579162	
galGal5	chr28:1281710-1282336(+)	H/ACA box	0.543749		Xenla9	chr7L:4156410-4156600(-)	H/ACA box	0.565685	
galGal5	chr5:15841244-15841474(+)	C/D box	0.54		Xenla9	chr7S:85805400-85805651(-)	H/ACA box	0.560260	
galGal5	chr8:5825119-5825693(+)	H/ACA box	0.529885		Xenla9	chr4S:20943430-20943860(+)	H/ACA box	0.557561	
galGal5	chr25:124602-125288(+)	H/ACA box	0.520570		Xenla9	chr1S:163788621-163789190(-)	C/D box	0.55	
					Xenla9	chr9_10S:44422890-44423280(+)	C/D box	0.54	
					Xenla9	chr9_10L:43867589-43868221(+)	H/ACA box	0.532362	
					Xenla9	chr7L:119915649-119915875(-)	H/ACA box	0.530115	
					Xenla9	chr9_10L:46262304-46262610(-)	H/ACA box	0.500000	

CHAPTER 4

7SL RNA independent of the signal recognition particle in red blood cells

INTRODUCTION

In a previous study, I reported the existence of cytoplasmic sisRNA in *Xenopus* oocytes (Chapter 1). To find cytoplasmic sisRNAs in other species, I analyzed mammalian red blood cells (RBCs), which lack nuclei (Chapter 2). During my analysis, I noticed that 7SL RNA was surprisingly abundant in RBCs.

7SL RNA is a major component of the eukaryotic signal recognition particle or SRP (Walter and Blobel, 1982). RNAs with a secondary structure similar to that of 7SL are found in SRPs from all domains of life (4.5S RNA in bacteria) (Struck *et al*., 1988). In addition to the core RNA subunit, the SRP contains six signal recognition proteins: SRP9/14 associated with the Alu domain of 7SL RNA and SRP 19/54/68/72 bound to the SRP 'specific' (S) domain of 7SL. The SRP mediates co-translational translocation of proteins to the ER (mechanism reviewed by Nagai *et al*. 2003). Therefore, the SRP is essential for cell survival, except for RBCs.

Erythroblasts have very condensed nuclei (10% the size of its original volume) and, as mentioned, in mammals (and some salamanders) RBCs extrude their nuclei before entering the circulation (Ji *et al*., 2011; Villolobos *et al*., 1988). During RBC maturation, organelles and ribosomes are partially or entirely discarded, presumably to increase the efficiency of gas exchange (Giger and Kalfa, 2015). RBCs remain stripped of their capacity to synthesize proteins throughout their lifespan (120 and 40 days in human and mouse, respectively). Therefore, it was a surprise to detect 7SL RNA in RBCs.

In this chapter, I report two additional unexpected findings: (1) a shorter transcript derived from the 7SL gene accumulates in mammalian RBCs. (2) 7SL RNA is associated with a new RNA-binding protein in RBCs.

RESULTS & DISCUSSION

7SL RNA in mature RBCs

In my previous study, I purified mouse and human RBCs by centrifugation on a Percoll gradient. The RBC RNA was purified and subjected to high-throughput sequencing. I noticed that a large fraction of reads mapped to the RN7SL gene, which encodes 7SL RNA. In the mouse sample, about a quarter of reads mapped to RN7SL without mismatch and, in human samples, about a third. The sequence and integrity of 7SL RNA in RBCs were confirmed by 5' and 3' RACE (data not shown). In comparable RNAseq datasets from cultured mouse and human cells, I found that about 10% of reads mapped to RN7SL (Fig. 19A). In contrast, other abundant ncRNAs such as 7SK (nuclear) and RNase MRP (nuclear and cytoplasmic) are 10 to 100 times less abundant in RBCs than in cell culture (Fig 19B). Hence, the enrichment for 7SL RNA in RBCs is likely due to the selective cellular RNA degradation occurring during RBC maturation (Burka, 1971) rather than higher transcription of 7SL in precursor cells.

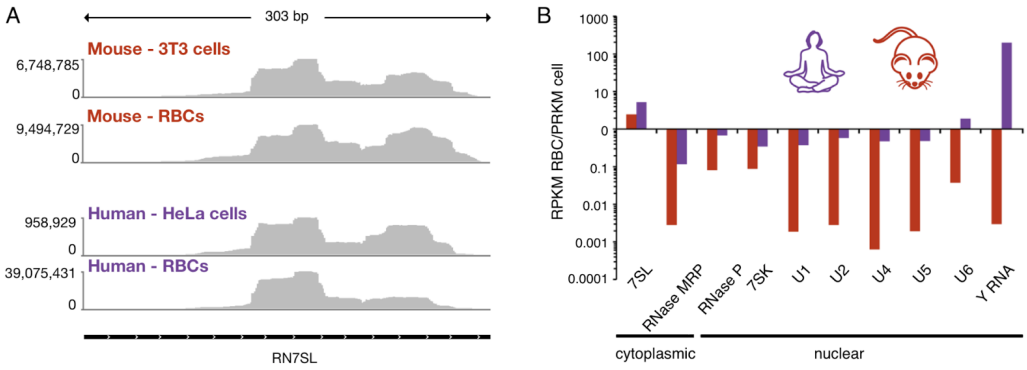


Figure 19. 7SL is enriched in RBCs relative to other ncRNAs. (A) Reads mapping to the RN7SL gene. A similar pattern is observed in RBCs and cultured cells (3T3 and HeLa cells). (B) The abundance of ncRNAs in RBCs relative to cultured mouse and human cells.

To examine the distribution of 7SL in RBCs more closely, Dr. Joseph Gall carried out single molecule in situ hybridization (smISH) on conventional blood smears from *Xenopus*, human and mouse. We were especially interested to determine whether 7SL was limited to a specific subset of cells, such as immature RBCs, or was present in all mature cells. Because 7SL is only 300 bases long, we used a mixture of three short probes, each of which consisted of 2 adjacent sequences of about 25 nucleotides (3 "ZZ" BaseScope probes from Advanced Cell Diagnostics, Inc (ACD)). RBCs exhibit a high level of autofluorescence throughout the visible spectrum, requiring the use of a chromogenic detection system (Fast Red). We found that RBCs of all three species are remarkably resistant to the hybridization procedure, when prefixed with ethanol and/or paraformaldehyde. Fig. 20A-D shows a field of RBCs from a blood smear of *X. tropicalis* prefixed in this way. Unknown to us before we began these experiments, a small number of RBCs lyse during the spreading procedure. Because *Xenopus* RBCs are nucleated, these lysed cells can still be recognized by the residual DAPI staining of their extruded nuclei (panel B). After hybridization, there is no signal in the intact RBCs, but strong hybridization in all cells that lysed during preparation of the smear (panels C and D).

Blood smears that were pretreated by heating to 60^{o} C for 1 hr, but not subjected to chemical fixation, showed signal in many intact RBCs (Fig. 20, panel E). The signal was still consistently higher in cells that lysed during the initial spreading procedure. In the case of human and mouse RBCs, which lack nuclei, the position of lysed cells may be inferred from the occasional "blobs" of heavy signal scattered among the intact RBCs (Fig. 20, panel F). The apparent resistance of RBCs to penetration by the probes remains unexplained and is in striking contrast to the ease with which cultured cells or cells in

tissue sections can be hybridized. Nevertheless, these ISH data strongly suggest that most, if not all, mature RBCs of human, mouse and *Xenopus* contain 7SL RNA, thereby directly confirming the biochemical data.

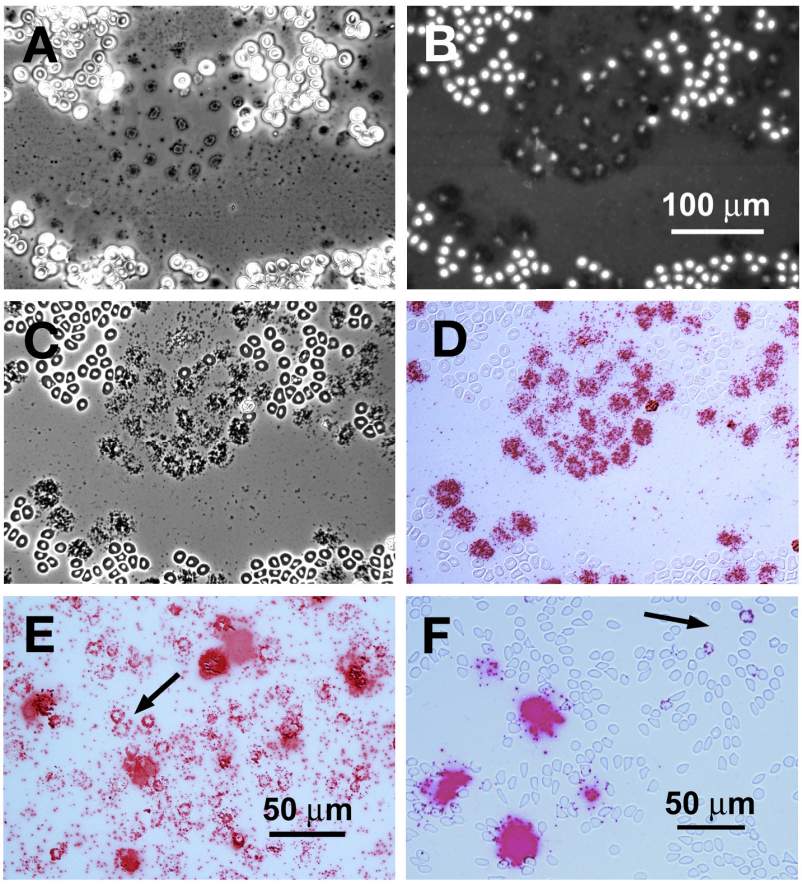


Figure 20. Single molecule in situ hybridization of 7SL RNA. (A-D) The same area of a blood smear of *Xenopus tropicalis*, prefixed with ethanol and 4% paraformaldehyde. (A) Phase contrast before any treatment (no mounting medium). A patch of lysed RBCs (black) surrounded by intact RBCs (white). (B-D) The same area after in situ hybridization with a single-molecule probe (red) against 7SL RNA. (B) Fluorescent DAPI stain (white) for DNA. Note that lysed RBCs are each identifiable by their DAPI stain, even though they have lost much of their DNA. The in situ hybridization label does not fluoresce. (C) Phase contrast image of cells in mounting medium under a coverslip. Most RBCs are still intact, showing the nucleus (white) and cytoplasm (black). (D) Brightfield image in which individual red dots presumably represent single 7SL molecules. Note intense label over each lysed RBC but complete absence of label over intact RBCs. (E) A different smear of *X. tropicalis* RBCs hybridized as in (A-D) but without prefixation. Some, but not all, intact RBCs (arrow) now show a few dots of label in the cytoplasm (arrow). The label over RBCs that lysed during spreading is so intense that individual molecules are not resolvable. (F) Hybridization of human RBCs with the same 7SL probe and no prefixation. A very few intact RBCs show label (arrow). The massive patches of label presumably represent single lysed RBCs or clusters of a few lysed cells. Because human RBCs lack nuclei, the identification of lysed cells is based on the similarity to the condition in *Xenopus* preparations.

Short ncRNA encoded by the RN7SL gene in mammalian RBCs

To further characterize 7SL in RBC, I electrophoresed mouse RBC RNA on an acrylamide gel and performed northern blotting against 7SL. The probe used is shown in red (Fig. 21A), highlighted on the secondary structure of 7SL (based on human 7SL, http://rth.dk/resources/rnp/SR PDB/). I detected a band at about 300 nt, confirming that 7SL RNA is stabilized as full-length in RBC in accordance with the 5' and 3' RACE results. Surprisingly, I detected an additional band that corresponds to a small ncRNA presumably derived from 7SL. This small ncRNA was readily detectable in mouse RBCs, but not in mouse brain, liver, or testes (Fig. 21B). There are presumably some RBCs in the latter three tissue samples, but at a too low concentration to detect the smaller band.

To determine the sequence of this short ncRNA, I made a small RNA library from gel-purified 60-80 nt long mouse RBC RNAs, and I cloned the library in *E. coli*. Colonies were screened by southern blotting, and positive ones were sequenced. Among the 11 sequenced colonies, 8 carried a plasmid containing the sequence between nucleotide 143 and 210 of 7SL (sequence highlighted in green, Fig. 21A). The three other colonies carried a plasmid with a shorter fragment, sequence between 143 and 207 of 7SL. I refer to this RNA as 'sRN7SL' for 'short 7SL RNA.' To confirm the sequence, I performed northern blotting with a probe against sRN7SL (highlighted in green, Fig. 21A) or against the downstream region (highlighted in blue, Fig. 21A). I detected the full-length 7SL RNA with both probes, but the smaller transcript reacted only to the sRN7SL probe (Fig. 21C). A ncRNA had been predicted at a similar region in the human RN7SL gene (ENST00000625125.2 - UCSC genome browser on GCRCh38). However, the prediction is linked to a piRNA, piR-45120 locus DQ577008, identified by Girard *et al*. (2006). The

sequence of sRN7SL overlaps that of piR-45120 by only six nucleotides, and sRN7SL predicted secondary structure differs from that of a pre-miRNA. Therefore, it is unlikely that sRN7SL is a precursor for a shorter RNA.

Because the 68 nt region of the RN7SL gene is relatively well conserved across vertebrates (Fig. 21D), I asked if sRN7SL was expressed in other species. I probed RNA from human, mouse, chicken, and *Xenopus* RBCs for sRN7SL with the full-length 7SL RNA probe. I detected full-length 7SL RNA in all samples, but the 68 nt band was present only in the human and mouse RBC samples (Fig. 21E).

One possibility is that the short transcript is a simple byproduct of 7SL RNA degradation. However, random byproducts typically run as a smear on acrylamide gels, do not have specific 5' and 3' ends, and would presumably be detectable in all RBCs. Instead, sRN7SL has distinct ends as determined by 5' and 3' RACE and is detectable only in mammals. Notably, I could not produce the sRN7SL from *in vitro* transcribed full-length 7SL added to mammalian RBC lysate (data not shown). Hence it is likely that sRN7SL is produced before RBC maturation.

7SL RNA binds to SART3 in RBCs

Detecting a short ncRNAs derived from 7SL was not the only surprise. I also found that 7SL forms a previously unknown RNP in RBCs. The human RBC proteome has been extensively studied by mass spectrometry (e.g., Kakhniashvili *et al*., 2004; Pasini *et al*., 2006; Roux-Dalvai *et al*., 2008; and Bell *et al.*, 2013; all summarized

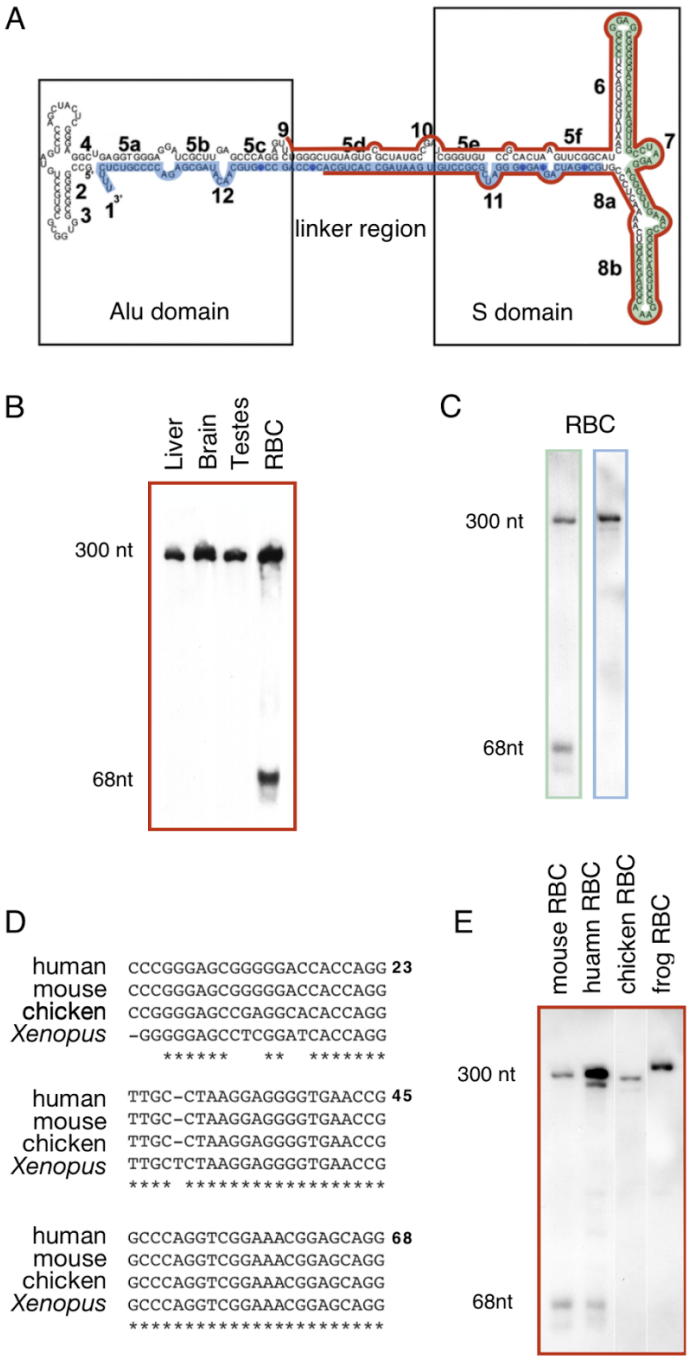


Figure 21. The RN7SL gene encodes 7SL RNA and a shorter ncRNA. (A) Human 7SL secondary structure: red, green and blue regions show the probe sequences used for the northern blots. (B) Northern blot analysis of 7SL using the red probe. A small band is detectable in the RBC sample only. (C) The sequence of sRN7SL is confirmed by northern blot analysis using the green (sRN7SL) and the blue regions (control). (D) Alignment of human, mouse, chicken and *Xenopus* partial 7SL sequence. (E) sRN7SL is detected by northern blot (red probe) only in mammalian (mouse and human) RBCs.

by Hegedüs *et al*. 2015). The majority of the SRP components were not detected, even in the most sensitive mass spectrometry experiments. Additionally, by fractionation of a rabbit reticulocyte extract, Avanesov (1988) identified 7SL RNA in a fraction containing mainly one protein, which had a M.W. different from that of the SRP proteins. Altogether, the evidence suggests that 7SL RNA in RBCs must be a component of a new RNP.

To determine the protein(s) bound to 7SL RNA in RBCs, I adapted the capture hybridization analysis of RNA targets (CHART) method (Simon *et al*., 2011). This method includes two steps. First, RNase H is used to identify regions of the target RNA that can be hybridized by antisense DNA oligos. This step takes advantage of the fact that RNase H cleaves RNA only when it is hybridized to DNA. Second, biotinylated oligos complementary to the determined 'open' regions are used to pull down the RNA and bound proteins.

I mixed RBC lysate with RNase H and a variety of DNA oligonucleotides. RNase H cut 7SL efficiently only in the presence of an oligonucleotide that targeted the linker domain (Fig. 22A). This result suggests that the linker domain may be free of protein. Interestingly, 7SL from a 3T3 cell lysate was not cut under the same conditions (Fig. 22B). This difference is consistent with 7SL RNA being in different complexes in the two cell types. I used this specific binding to purify the 7SL complex from RBCs.

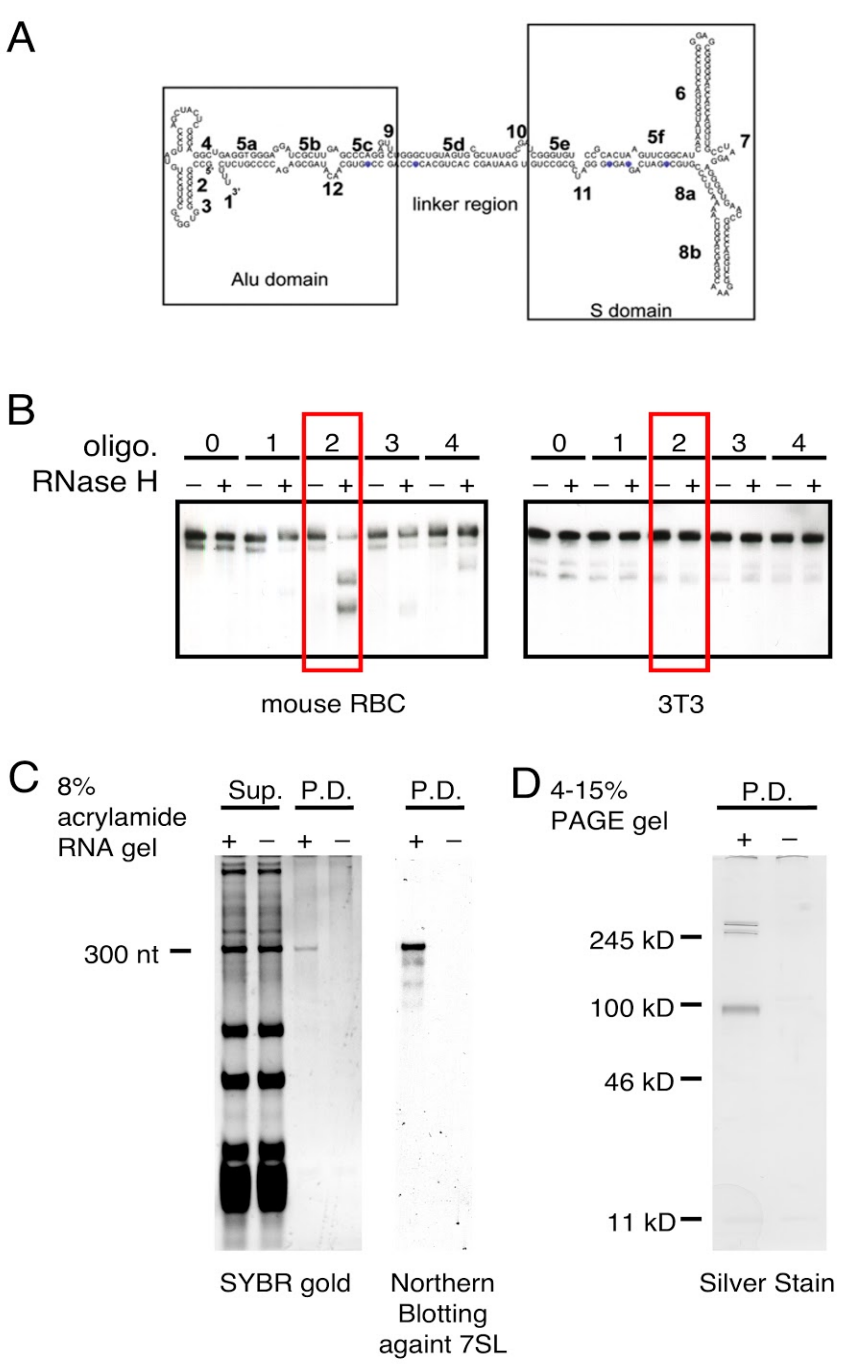


Figure 22. Affinity purification of 7SL RNA. (A) Sequence of oligos used to probe 7SL in RBCs. (B) Cleavage of 7SL by RNase H using the different oligos. In an RBC lysate RNase H cleaved 7SL only with oligo 2 (red). (C) Total RNA detected by SYBR gold in the supernatant (Sup.) and the pull down (P.D.) when using 7SL antisense probe (+) and control (-). Analysis of the pulled down RNA by northern blot showed readily detectable 7SL. (D) Total proteins pulled down with 7SL. The major bands were identified as proteins of the cytoskeleton.

For the affinity purification of 7SL RNA, I used a 'linker targeting' antisense oligo with methylated nucleotides and a desthiobiotin-conjugated 3'end nucleotide. As a negative control, I used a desthiobiotin-conjugated oligo antisense to a *Xenopus tropicalis* specific RNA. Pulled-down RNA was run on an acrylamide gel and stained with SYBR-gold. The target was detected in the pull-down but not in the control. Importantly, 7SL was the major RNA pulled-down, suggesting that sRN7SL is not associated with 7SL (Fig. 22C). Next, I analyzed the proteins by mass spectrometry and found enrichment for many actin-associated and membrane proteins (by at least 3 fold) in the pull-down compared to control, suggesting that 7SL is associated with a membrane–cytoskeleton complex (Fig. 22D). Our observation is consistent with an earlier suggestion that RNAs in erythroid cells gain resistance to endogenous nucleases by binding to the membrane (Burka *et al.*, 1967). Of particular interest, I detected one RNA-binding protein: squamous cell carcinoma antigen recognized by T-cells 3 or SART3 (Fig. 23).

SART3 has been shown to bind to tubulin and actin (Timani *et al.*, 2013); therefore, it may be anchoring 7SL RNA to the cytoskeleton. Additionally, SART3 has been shown to have many functions, including ubiquitin-mediated regulation of proteins and oxygen sensing (Whitmill *et al.*, 2016). Both SART3 and 7SL RNA have been implicated in p53 mRNA homeostasis (Abdelmohsen *et al.*, 2014; Timani *et al.*, 2015). Therefore, it is possible that both work together in RBCs. For instance, 7SL could act as a transient scaffold for recruiting proteins to SART3 and the membrane-cytoskeleton complex in general. This system would be advantageous for rapid responses to external cues or local damage and aid in a rapid immune response.

.

ACCESSION	GENE NAME	7SL #peptide	control #peptide	comments
NP_068683.1	tropomodulin-1	52	0	
NP_058622.1	squamous cell carcinoma antigen recognized by T-cells 3	31	0	RNA-binding
NP_001186044.1	equilibrative nucleoside transporter 1 isoform 2	28	0	
NP_001035916.1	alpha-synuclein	28	0	
NP_033931.4	carbonic anhydrase 2	19	0	
NP_034499.3	glycophorin-A	18	0	
NP_032405.3	importin subunit beta-1	17	0	RNA-binding
NP_001153027.1	platelet glycoprotein 4	15	0	
NP_032647.1	55 kDa erythrocyte membrane protein	15	0	
NP_001239591.1	dematin isoform 1	53	10	
NP_001188320.1	hemoglobin, beta adult s chain	120	27	
NP_001104253.1	ankyrin-1 isoform 1	790	196	
NP_001258786.1	alpha-adducin isoform 1	70	19	
NP_001265090.1	hemoglobin subunit beta-1	144	40	
NP_058652.1	hemoglobin subunit beta-2	105	32	
NP_001240667.1	tropomyosin	167	57	
NP_001161421.1	plasma membrane calcium ATPase 4 isoform a	27	9	
NP_031498.1	aquaporin-1	60	21	
NP_038703.3	spectrin beta chain, erythrocytic	1174	428	

legend	cytoskeleton
	hemoglobin
	plasma membrane

Figure 23. Proteins enriched by at least 3 fold in the pull down fraction compared to control. Cytoskeletal proteins are marked in green, hemoglobins in red and plasma membrane proteins in blue.

Other ncRNAs detected in mature RBCs

On a side note, many other ncRNAs were enriched in RBCs relative to other somatic cells. For instance, in human RBCs but not in mouse, I detected unexpectedly high levels of the snRNA U6, which typically accumulates in the nucleus, where it facilitates mRNA splicing (Fig. 19). Although the half-life of U6 is shorter than that of other snRNAs (Fury MG and Zieve GW,1996), it was the most enriched snRNA in RBCs. Known U6-binding proteins, LSM proteins, were detected in the cytosol of RBCs (Hegedűs *et al*., 2015). The U6 RNP may carry out a new function in the cytoplasm. Similarly, Y RNAs and the Y RNA-binding protein Ro60, were detected in human but not in mouse RBCs (RNA: Fig 19) (proteins were detected by mass spectrometry:

Hegedűs *et al*., 2015). In the nucleus Y RNAs may regulate DNA replication, whereas their function in the cytoplasm remains largely unknown (Christov *et al.*, 2006).

Human RBCs have a longer lifespan than mouse RBCs (120 d vs 40 d). One intriguing possibility is that RBCs may take advantage of ncRNAs to respond to oxidative stress, extreme cytoskeleton remodeling, and external cues. In other words, RNAs could be repurposed in different conditions to extend the lifespan of RBCs.

MATERIALS AND METHODS

Blood smear

A drop of whole blood was spread quickly on a microscope slide and dried overnight or longer. In some cases, the smear was fixed in 100% ethanol and/or 4% paraformaldehyde.

Single molecule in situ hybridization

Experiments were carried by Dr. Joseph G. Gall according to the manufacturer's protocols (RNAScope and BaseScope probes from ACDBio)

RNA purification

RNA was extracted with TRIzol reagent (Ambion) and purified with the Direct-zol RNA MiniPrep kit (Zymo Research). DNase I treatment was performed on the Direct-zol column. RNA was quantitated with a Nanodrop One (Thermo Scientific) and a Bioanalyzer 2100 (Agilent).

Northern blots

Up to 1 μg of RNA per sample was separated on an 8% polyacrylamide–8 M urea gel and transferred onto a nylon membrane (Zeta Probe GT, Bio-Rad). RNA was probed with dsDNA labeled with digoxigenin (DIG)-dUTP in hybridization buffer (Roche) and detected using an anti-DIG antibody conjugated with alkaline phosphatase (Roche) and CDP-*Star* chemiluminescent substrate (Roche).

RNase H

Cells were homogenized in lysis buffer (50 mM hepes pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.2 U RNasin, one protease inhibitor cocktail tablet (cOmplete Tablets - Mini EASYpack, Roche, and 1% Tween-20). 20 µl of lysate was incubated 1 h at 37°C with 1 µL of RNase H (20 U) and 1mM (final) antisense oligo prior to RNA purification.

RACE/small RNA library

RNA was run on an 8% polyacrylamide–8 M urea gel. RNAs ranging from 200-400nt and 50-100 nt were cut and extracted from the gel in extraction buffer (300 mM NaOAc, 1mM EDTA, 0.25% SDS) at room temperature overnight. Oligos were ligated to the 5' and 3'end according to the NEBNext multiplex small RNA library protocol. The RNA was converted to cDNA with an oligo targeting the 3'end ligated sequence. It was then amplified either with NEBnext oligos to make a library or with an internal oligo to sequence the end of 7SL specifically.

RNA pull-down assay

100 µL of packed mouse RBCs were homogenized in lysis buffer and incubated 1 h at 37°C with 5 µL of 10 µM biotinylated antisense oligonucleotide (purchased from Integrated DNA Technologies):
/5BioTinTEG/iSp18/mAmUmCmGmGmCmAmUmAmGmCmGmCmAmCmUmAmCmAmGmCmCmCmAmGmAmAmCmUmCmCmUmGmGmAmCmUmCmAmAmG/iSp18/3BioTEG/).

Washed Dynabeads MyOne streptavidin T1 (Thermo Fisher Scientific) were added (100 uL) and the mix was further incubated at 37°C for 1 h. RNA was purified from beads that were directly added to TRIzol, run on an 8% acrylamide 8M urea gel, and analyzed by SYBR-gold and northern blotting. The pulled-down complexes were eluted (after five 10 min washes in lysis buffer) with 50 µL of nanopure water by heating gradually (0.5°C per sec) to 70°C (based on Holmberg *et al*., 2005). 5µL of the eluate was run on a 4-15% polyacrylamide gel (mini-protean TGX precast protein gels from BioRad) and silver stained (Pierce Silver Stain for Mass Spectrometry from Thermo Fisher Scientific) according to the manufacturer's protocol.

Mass Spectrometry

Mass Spectrometry was carried at Johns Hopkins Medical School's Mass Spectrometry and Proteomics Facility. There the samples were digested with trypsin, desalted, and analyzed by liquid chromatography tandem-mass spectrometry on a QExactive plus instrument (Thermo Fisher).

REFERENCES

Avanesov A.Ts. **1988**. Isolation and characterization of a new class of cytoplasmic RNP particles containing 7SL RNA. Biokhimiya 53(11): 1912-1919.

Abdelmohsen K, Panda AC, Kang MJ, Guo R, Kim J, Grammatikakis I, Yoon JH, Dudekula DB, Noh JH, Yang X, Martindale JL and Gorospe M. **2014** 7SL RNA represses p53 translation by competing with HuR. *Nucleic Acids Res*. **42**:10099–10111.

Bell AJ, Satchwell TJ, Heesom KJ, Hawley BR, Kupzig S, Hazell M, Mushens R, Herman A and Toye AM. **2013**. Protein Distribution during Human Erythroblast Enucleation *In Vitro*. *PLoS ONE* **8**(4): e60300.

Burka ER. **1971**. Erythroid cell RNase: activation by urea and localization to the cell membrane. *J Clin Invest*. **50**(1):60-8.

Burka ER, Schreml W, and Kick CJ. **1967.** Membrane-Bound Ribonucleic Acid in Mammalian Erythroid Cells. Biochemistry *6* (9), 2840-2847

Christov, C. P., Gardiner, T. J., Szüts, D., & Krude, T. (2006). Functional Requirement of Noncoding Y RNAs for Human Chromosomal DNA Replication. Molecular and Cellular Biology, 26(18), 6993–7004.

Fury MG and Zieve GW. **1996.** U6 snRNA maturation and stability. *Exp Cell Res*, 228,160–163.

Giger KM, Kalfa TA. **2015**. Phylogenetic and Ontogenetic View of Erythroblastic Islands. *BioMed Research International*. **2015**:873628.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. **2006**. A germline-specific class of small RNAs binds mammalian piwi proteins. *Nature* 442:199–202.

Hegedűs, T., Chaubey, P.M., Várady, G., Szabó, E., Sarankó, H., Hofstetter, L., Roschitzki, B., Stieger, B. and Sarkadi, B. **2015.** Inconsistencies in the red blood cell membrane proteome analysis: generation of a database for research and diagnostic applications. *Database* 2015: bav056.

Holmberg A, Blomstergren A, Nord O, Lukacs M, Lundeberg J., Uhlén M. **2005** The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis.* **26**:501–510.

Ji P, Murata-Hori M, Lodish HF. **2011**. Formation of mammalian erythrocytes: Chromatin condensation and enucleation. *Trends in cell biology*. **21**(7):409-415.

Kakhniashvili, D. G., Bulla, L. A., Jr., and Goodman, S. R. **2004**. The human erythrocyte proteome: Analysis by ion trap mass spectrometry. *Mol. Cell. Proteomics* **3**, 501–509

Nagai K, Oubridge C, Kuglstatter A, Menichelli E, Isel C and Jovine L. **2003**. Structure, function and evolution of the signal recognition particle. *The EMBO Journal* **22**(14):3479-3485.

Pasini EM, Kirkegaard M, Mortensen P, Lutz HU, Thomas AW and Mann M. **2006**. In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood.* **108**:791-801.

Roux-Dalvai F, Gonzalez de Peredo A, Simó C, Guerrier L, Bouyssié D, Zanella A, Citterio A, Burlet-Schiltz O, Boschetti E, Righetti PG and Monsarrat B. **2008**. Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. *Mol. Cell. Proteomics* **7**, 2254–2269

Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE. **2011**. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci*. **108**:20497–20502.

Struck JCR, Toschka HY, Specht T and Erdmann VA. **1988**. Common structural features between eukaryotic 7SL RNAs, eubacterial 4.5S RNA and scRNA and archaebacterial 7S RNA. *Nucleic Acids Research.* **16**(15):7740.

Timani KA, Liu Y, He JJ. **2013**. Tip110 interacts with YB-1 and regulates each other's function. BMC Mol Biol 14:14. 10.1186/1471-2199-14-14.

Timani KA, Liu Y, Fan Y. , Mohammad K.S. , He JJ. **2015**. Tip110 regulates the cross talk between p53 and hypoxia-inducible factor 1α under hypoxia and promotes survival of cancer cells. Mol. Cell. Biol., 35, pp. 2254-2264

Villolobos M, Leon P, Sessions SK and Kezer J. **1988**. Enucleated erythrocytes in plethodontid salamanders. *Herpetologica*. **44**(2):243–250.

Walter P. and Blobel G. **1982**. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature*. **299**:691–698.

Whitmill A, Timani KA, Liu Y, He JJ. **2016**. Tip110: Physical properties, primary structure, and biological functions. *Life Sci.* 149, 79.

DISCUSSION

What we know about introns, then and now

Introns represent a large fraction of most eukaryotic genomes and they continue to surprise us. Since their discovery (Berget et al., 1977; Chow et al., 1977), introns have been linked to essential molecular functions such as transcription (Vasil et al., 1989), translation (Moore, 2005), and mRNA surveillance (Maquat, 2004). In short, some introns encode regulatory elements that can enhance or silence transcription. After transcription, splicing of intronic RNA results in the deposition of the exon junction complex, which facilitates translation and mRNA surveillance (reviewed by Chorev and Carmel, 2012). Finally, intronic RNAs can be retained in mRNA, enhancing protein diversity, or they can encode ncRNAs such as micro (mi)RNA, small nucleolar (sno)RNA and long non-coding (lnc)RNA.

In my thesis, I report another layer of complexity to intronic RNAs. Whereas these RNAs are usually linearized and degraded within minutes, I established that some escape this fate by remaining circular. These circular sisRNAs are found in various cultured cells and tissues in humans, mice, chickens, frogs, and zebrafish. Most abundant stable lariats are exported to the cytoplasm of these cells by the NXF1 pathway.

Although the exact mechanism remains unknown, I propose two aspects of sisRNAs that affect their stability. First, export to the cytoplasm can be sufficient to stabilize lariats (Chapter 1 and 3). Second, most circular sisRNAs have a C branch point (except those in *Xenopus*) that may provide stability, since the debranching enzyme does not linearize C-branched lariats efficiently (Chapter 2).

Why were cytoplasmic sisRNAs never detected before?

I presented evidence that sisRNAs could be detected easily by high-throughput sequencing, yet very few labs reported their existence (Gardner *et al*., 2012; Zhang *et al.*, 2013; Talhouarne and Gall, 2014; Li *et al*., 2016). Detection of sisRNAs requires specific features of RNAseq experimental design and analysis.

Most RNAseq experiments are carried out after poly (A) selection to increase coverage of processed exonic transcripts. This method is advantageous for mRNA analysis but selects against any circular RNA.

Bioinformatic analysis of circular RNA can be challenging. There are several methods available, all of which show different biases (review by Szabo and Salzman, 2016). Even fewer bioinformatic tools are designed for detecting and quantifying intronic reads.

Finally, when intronic reads *are* detected, they are often discarded, because they are thought to represent contamination from pre-mRNA. Assumptions are as accurate as our understanding of the model system we study. We discovered nuclear and cytoplasmic sisRNAs because we could get a significant amount of pure stable nuclear and cytoplasmic RNA. Our example highlights the importance of choosing the right model system.

Potential function of cytoplasmic sisRNAs

Several studies suggest that nuclear sisRNAs can regulate their host gene expression. In cell culture, nuclear knock-down of a circular sisRNA led to decrease of the corresponding mRNA (Zhang et al., 2013). Similarly, in *Drosophila* embryos, genetic knock-out of maternally deposited intronic RNA led to down-regulation of host gene

zygotic transcription (Tay and Pek, 2017). Here I propose that cytoplasmic sisRNAs do not generally function in *cis*.

In Chapter 1, I analyzed cytoplasmic sisRNAs from *Xenopus tropicalis* and *Xenopus laevis* oocytes. Both amphibian oocytes share a similar growth program; for instance, 5S RNA is expressed at an early stage of oocyte growth in both species. If cytoplasmic sisRNAs facilitated gene regulation in cis, we would expect that homologs would encode the same cytoplasmic sisRNA in both species. However, I found that most of the genes hosting a cytoplasmic sisRNA in *X. tropicalis* were different from those in *X. laevis*. And when sisRNAs were encoded by homologous genes in both species, the stable lariats were made from different introns.

In Chapter 2, I report that sisRNAs, but not their cognate mRNA, persist in circulating human and mouse RBCs. As previously mentioned, these cells lack nuclei and cytoplasmic organelles, including ribosomes. Therefore, RBC sisRNAs cannot be regulating host genes or cognate mRNAs. In the oocyte, Dr. Joseph Gall demonstrated that cytoplasmic sisRNAs did not co-localize with their cognate mRNA, which suggested that no stable complex formed between these transcripts.

In Chapter 3, I showed that dyskerin can bind sisRNAs that have a snoRNA motif. Also, Li *et al*. (2017) reported that dicer could bind certain lariats in plants. One possibility is that sisRNAs are sponges for dyskerin and dicer. In other words, sisRNAs can accumulate and sequester the two nuclear proteins away from their canonical targets. It is likely that we will find more RNA-binding proteins that can be sequestered by sisRNAs.

Non-coding RNAs can have multiple functions as demonstrated by my work on 7SL RNA in RBCs (Chapter 4) and many other studies (including Bardou *et al.,* 2011; Buck and Griffiths, 1982; Falaleeva *et al.,* 2017; Gimpel and Brantl, 2017; Green *et al.,* 2010; Holey *et al.,* 2015; Jobert *et al.,* 2009; Katz *et al*., 2016; Kumari and Sampath *et al.,* 2015; Liu *et al.*, 2015; Martens-Uzunova *et al.,* 2013; Zhang *et al*., 2013; Zhao *et al*., 2007). It is possible that sisRNAs shuttle between the nucleus and the cytoplasm to regulate *both* transcription and the availability of specific proteins.

REFERENCES

Bardou F., Merchan F., Ariel F., Crespi M. **2011**. Dual RNAs in plants. Biochimie. 93:1950–1954.

Berget SM, Moore C, Sharp PA. **1977**. Spliced segments at the 59 terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci 74: 3171–3175.

Buck, M., and Griffiths, E. **1982**. Iron mediated methylthidation of tRNA as a regulator of operon expression in Escherichia co/i. Nucl. Acids Res. 10, 2609-2624.

Chorev M and Carmel L. **2012**. The Function of Introns. Front Genet. 3: 55.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. **1977**. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12: 1–8.

Falaleeva M, Welden JR, Duncan MC, Stamm S. **2017**. C/D-box snoRNAs form methylating and non methylating ribonucleoprotein complexes: old dogs show new tricks. Bioessays. 39(6).

Gardner EJ, Nizami ZF, Talbot CC, Jr., Gall JG. **2012**. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of Xenopus tropicalis. Genes Dev 26: 2550-2559.

Gimpel, M. and Brantl, S. **2017**, Dual-function small regulatory RNAs in bacteria. Molecular Microbiology, 103: 387–397.

Green NJ, Grundy FJ, Henkin TM. **2010**. The T box mechanism: tRNA as a regulatory molecule. FEBS Lett 584: 318-324.

Holley CL, Li MW, Scruggs BS, Matkovich SJ, Ory DS, Schaffer JE. **2015**. Cytosolic accumulation of small nucleolar RNAs (snoRNAs) is dynamically regulated by NADPH oxidase. J Biol Chem. 290(18):11741–11748.

Jobert L, Pinzón N, Van Herreweghe E, Jády BE, Guialis A, Kiss T, Tora L. **2009**. Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. EMBO Rep. 10:494-500.

Katz A., Elgamal S., Rajkovic A., Ibba M. **2016**. Non-canonical roles of tRNAs and tRNA mimics in bacterial cell biology. Mol. Microbiol. 101:545–548.

Kumari P and Sampath K. **2015**. cncRNAs: Bi-functional RNAs with protein coding and non-coding functions. Semin Cell Dev Biol. 47-48:40-51.

Maquat LE. **2004**. Nonsense mediated mRNA decay: splicing, translation and mRNP dynamics. Nat. Rev. Mol. Cell. Biol. 5, 89–99.

Martens-Uzunova ES, Olvedy M and Jenster G. **2013**. Beyond microRNA – Novel RNAs derived from small non-coding RNA and their implication in cancer. Cancer Lett. 1;340(2):201-11.

Moore MJ. **2005**. From birth to death: the complex lives of eukaryotic mRNAs. Science 309, 1514–1518

Li Z, Wang S, Cheng J, Su C, Zhong S, Liu Q, Yuda Fang Y, Yu Y, Lv H, Zheng Y and Zheng B. **2016**. Intron Lariat RNA Inhibits MicroRNA Biogenesis by Sequestering the Dicing Complex in *Arabidopsis.* PLoS Genet. 12(11): e1006422.

Liu N, Dai Q, Zheng G, He C, Parisien M, and Pan T. **2015**. N^6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* 518, 560–564.

Szabo L and Salzman J. **2016.** Detecting circular RNAs: bioinformatic and experimental challenges. *Nature Reviews Genetics.* 17, 679–692.

Talhouarne GJS and Gall JG. **2014**. Lariat intronic RNAs in the cytoplasm of Xenopus tropicalis oocytes. RNA. 20:1476–87.

Tay ML, Pek JW. **2017**. Maternally Inherited Stable Intronic Sequence RNA Triggers a Self-Reinforcing Feedback Loop during Development. Curr Biol. 3;27(7):1062-1067.

Vasil V, Clancy M, Ferl RJ, Vasil IK, and Hannah LC. **1989**. Increased gene expression by the first intron of maize shrunken-1 locus in grass species. Plant Physiol. 91, 1575–1579.

Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. **2013**. Circular intronic long noncoding RNAs. Mol Cell 51: 792-806.

Zhang, Z., Zhu, Z., Watabe, K., Zhang, X., Bai, C., Xu, M., Wu, F., and Mo, Y. **2013** Negative regulation of lncRNA GAS5 by miR-21. Cell Death & Differentiation 20, 1558-1568

Zhao X, Patton JR, Ghosh SK, Fischel-Ghodsian N, Shen L, and Spanjaard RA. **2007**. Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling. Mol Endocrinol 21:686–699.

APPENDIX

Table S1. Primers for chapter 1

EXPERIMENTS		Orientation	*arfgap2(1)*	*arfgap2(2)*
detection	mRNA	forward	AACCCGCATCTACCACAGTTACCA	CGCCAAAGCCATTTCCTCAGACAT
		reverse	AAGTTCCTGATACGCCAGCCTCAT	TCACTCCATTAGCCAGAACTGCCA
	sisRNA	forward	TACGCGGGCAGGGACTAATAACAA	CAGCCAGCCAGCCAATTATTTCCA
		reverse	AGAGTGCACACTTGCTCTCCTTGA	TTCATCAAGAGGTCTAGGGCAGCA
in vitro transcription construct	pre-mRNA	forward	AACCCGCATCTACCACAGTTACCA	CGCCAAAGCCATTTCCTCAGACAT
		reverse	AAGTTCCTGATACGCCAGCCTCAT	TCACTCCATTAGCCAGAACTGCCA
	sisRNA	forward	ATTAACCCTCACTAAAGGAACTGGAGCTGGTTCCAACTGAGA	ATTAACCCTCACTAAAGGAGACGCCAGTTGCTAACAAGTCCT
		reverse	GGATCCAGGACTTGTTAGCAACTGGCGTCT	GGATCCAGCCGGCCATATACGCAATTTCAC
lariat detection	sisRNA	forward	CCTTTCAGTTGAGGAGGCGA	CCCCGCTAACTTGACCCATT
		reverse	GGACTTCAGGACACTCGCAA	GGCAGCAGGGAAAACCACTA
inverted ivt contruct	sisRNA-piece1	forward	TAATACGACTCACTATAGGGGCAGCCATGTGATGCCCTATA	TAATACGACTCACTATAGGGAAATGTAACCTGCTGTGGGTCGC
		reverse	GAAATGGACCCACTCATGAGCAAGTGGTGCCCCAAATG	GCCGTTCTACTGACTCACTTGTGCCCAAACATGCAAATCA
	sisRNA-piece2	forward	GGGCACCACTTGCTCATGAGTGGGTCCATTTCCATTCTGTATA	GTTTGGGCACAAGTGAGTCAGTAGAACGGCCAACTGG
		reverse	GTAGGGCTCAGACTGAAGGC	AACGGAAGAGATCGGTGGCCA

EXPERIMENTS		Orientation	*eif4a1*	*faf2*
detection	mRNA	forward	AGATTACATGGGTGCCTCTTGCCA	TCGAATCAATGGACCAATGTCGCC
		reverse	TCAACATCTCGTCGGCTTCATCCA	GCTGTTAAGGAAATGGGTGACCTC
	sisRNA	forward	TTGGAATGAGGCTGTTTATGGGCG	ATTGATTAACCAGCAGCCTGAAGA
		reverse	ACTCTGGCCCACAGAGCTTACAAT	TATGCTCAGACCTGCTGTTGCACT
in vitro transcription construct	pre-mRNA	forward	AGATTACATGGGTGCCTCTTGCCA	TATTAACCCTCACTAAAGGACTTGACAGGTATCGAATCAA
		reverse	TCAACATCTCGTCGGCTTCATCCA	TTTTTTTTTTTTTTTTTTTTCCTTGGTTGTGGCCGAGACACCACATA
	sisRNA	forward	ATTAACCCTCACTAAAGGATCCAGCAGCGAGCTATTATGCCT	ATTAACCCTCACTAAAGGGGACGAACGTTCCTTTGCTTCTGA
		reverse	GGATCCTGAACTTCTTGGTCACCTCGAGCA	GGATCCTTTCTTGGTCAGCACGCAGAGATG
lariat detection	sisRNA	forward	CTTATATCTGGCGGGGGCAG	AGTGCAACAGCAGGTCTGAGCATA
		reverse	GTACCGCCCATAAACAGCCT	TCTTCAGGCTGCTGGTTAATCAAT
inverted ivt contruct	sisRNA-piece1	forward	TAATACGACTCACTATAGGGGAGGAATCTGCCATAGAAGTGCCT	ATTAACCCTCACTAAAGGGAACGTACCTTTACCTTTCTCTGTTAA
		reverse	GCTCATGATACTCACTGAGTGCTGCACTGGCGCCACG	GGTTCAGGTACATACTCATCAGAGAACAGTGTACTCTGTA
	sisRNA-piece2	forward	CCAGTGCAGCACTCAGTGAGTATCATGAGCCTGCCTCTG	AGTACACTGTTCTCTGATGAGTATGTACCTGAACCAGCAGGGAA
		reverse	TTTGCTGATGGGTAGTGAGCAT	ACACACAGTTAGTCATATTGCAAA

Table S2. Primers for chapter 2

target	orientation	sequence	experiment
pnpo2 mRNA	fwd	GCG TGT GGA AGG CCC TGT GAA GAA A	RT-PCR
pnpo2 mRNA	rev	TAC CTC TTG GTC CTG GTA GA	RT-PCR
pnpo2 intron, inward	fwd	GAG ATG CCA GGG TCT TGT CAA A	RT-PCR
pnpo2 intron, inward	rev	GGG ACA GAG GTG AGG GAA GTA	RT-PCR
pnpo2 intron, outward	fwd	GAG CCT AGG TTT GTA AGT CCT TAA	RT-PCR
pnpo2 intron, outward	rev	AGA CAG CCT CTG GGG ATG AA	RT-PCR
piezo mRNA	fwd	TTC GTT CGG CGC TTG CTA GAA	RT-PCR
piezo mRNA	rev	GGA GTA CTC ATG CGG GTT GAC AA	RT-PCR
piezo intron, inward	fwd	TGG GCT GTT TCC TCC TCT GAA	RT-PCR
piezo intron, inward	rev	ACC AGG CAG CAT CGG GAC TAG A	RT-PCR
piezo intron, outward	fwd	CAG TCG TCC TGG AGT CTA GT	RT-PCR
piezo intron, outward	rev	CTG CTT CAG AGG AGG AAA CA	RT-PCR
tnfrsf14 mRNA	fwd	CCC AGG CTA CTT CTG TGA GAA	RT-PCR
tnfrsf14 mRNA	rev	CCT GAG TCC CTC CAA GTG AGA A	RT-PCR
tnfrsf14 intron, inward	fwd	CCT CCA GCT TCA GGA TTT CAG AAA	RT-PCR
tnfrsf14 intron, inward	rev	GCA TTG GAG CTG AGG TGG AAT A	RT-PCR
tnfrsf14 intron, outward	fwd	TGG GTT GTG AGG CTC CAG AA	RT-PCR
tnfrsf14 intron, outward	rev	CAG CGG GCT TAC TGG AAC AA	RT-PCR
add1 mRNA	fwd	GGT CCA GGA GAT GAG GAA CAA	RT-PCR
add1 mRNA	rev	GTC CAT CAT CAC ACC ACA CAA	RT-PCR
add1 intron, inward	fwd	AGC CAC GAA CAG TCA AGA GTA G	RT-PCR
add1 intron, inward	rev	ACA CAC CCA CTT TAG GGT ACT G	RT-PCR
add1 intron, outward	fwd	ATA TGG CCG GGG TCT TAA G	RT-PCR
add1 intron, outward	rev	CCC CTC CCA AAG CTA TAT A	RT-PCR
faf2 intron inward/outward	all	published (Talhouarne and Gall, 2014)	RT-PCR
hcfc	fwd	TAG CCT TTG GTG AAT GCC ACT	northern blot probe
hcfc	rev	ATG GCC TGG GCT AGA ACT AT	northern blot probe
dut	fwd	TGT GTG TCC AGT GGA GGA	northern blot probe
dut	rev	GGA AAG CCA ATC AAG TCT CAA AC	northern blot probe
faf2	fwd	TAA TAC GAC TCA CTA TAG GGA TGC CTC CTA CAG GAC TTG ACA	northern blot probe
faf2	rev	CCT GGC TGT ATG TAC CCT GGT AAA	northern blot probe
mCherry	fwd	GGG CGA GGA GGA CAA CAT GGC CAT CAT CAA	northern blot probe
mCherry	rev	GCC GCG CAG CTT CAC CTT GTA GAT GAA	northern blot probe

Table S3. Primers for chapter 3. (red boxes represent primers designed by Svetlana Deryusheva)

primer description (species, orientation, target)	primer sequence	primer description	primer sequence
human - fwd - inward / slb-snoRD46	GTGACAGTTGTGCGTTCTTTC	chicken - fwd - inward / ci-rp11	GCATTCTTCCCTGGTGACTT
human - rev - inward / slb-snoRD46	AACAAGACTAGGCTTCTGACTATG	chicken - rev - inward / ci-rp11	GACAACCTCAGAGCCAGTTATT
human - fwd - outward / slb-snoRD46	CATAGTCAGAAGCCTAGTCTTGTT	chicken - fwd - outward / ci-rp11	AATAACTGGCTCTGAGGTTGTC
human - rev - outward / slb-snoRD46	GAAAGAACGCACAACTGTCAC	chicken - rev - outward / ci-rp11	AAGTCACCAGGGAAGAATGC
human - fwd - inward / ci-nisch	GAGCTTGGAGTGTGTGGTT	chicken - fwd - inward / slb-snoRD14	TAGATCTGCAGGAGACCTTCTT
human - rev - inward / ci-nisch	CTTCTCATCACTGGCTGTTC	chicken - rev - inward / slb-snoRD14	CAACACTCGTCTTTCGACTTCT
human - fwd - outward / ci-nisch	GAACAGCCAGTGATGAGAAG	chicken - fwd - outward / slb-snoRD14	AGAAGTCGAAAGACGAGTGTTG
human - rev - outward / ci-nisch	AACCACACACTCCAAGCTC	chicken - rev - outward / slb-snoRD14	AAGAAGGTCTCCTGCAGATCTA
mouse - fwd - inward / slb-snoRD57	TTGGTTGTGGTAGCTGAGTG	X laevis - fwd - inward / slb-snoRD33	TTGTTAGCAGAGGTCAGTGATG
mouse - rev - inward / slb-snoRD57	CTTGCTGGATCAGGCTCATTA	X laevis - rev - inward / slb-snoRD33	CAACGTACTTAGGTCTCAGATGG
mouse - fwd - outward / slb-snoRD57	TAATGAGCCTGATCCAGCAAG	X laevis - fwd - outward / slb-snoRD33	CCATCTGAGACCTAAGTACGTTG
mouse - rev - outward / slb-snoRD57	CACTCAGCTACCACAACCAA	X laevis - rev - outward / slb-snoRD33	CATCACTGACCTCTGCTAACAA
mouse - fwd - inward / ci-dut	TGTGTGTCCAGTGGAGGA	X laevis - fwd - inward / slb-snoRD56	CACGTTCCATCTGCCAGT
mouse - rev - inward / ci-dut	GGAAAGCCAATCAAGTCTCAAAC	X laevis - rev - inward / slb-snoRD56	ACACGCAGCACGTGATTA
mouse - fwd - outward / ci-dut	GTTTGAGACTTGATTGGCTTTCC	X laevis - fwd - outward / slb-snoRD56	TAATCACGTGCTGCGTGT
mouse - rev - outward / ci-dut	TCCTCCACTGGACACACA	X laevis - rev - outward / slb-snoRD56	ACTGGCAGATGGAACGTG
Yeast U2 snRNA	[6~ FAM]GGGTGCCAAAAAATGTGTATTGTAACA		
Yeast 18S rRNA (SnR54)	[6~ FAM]CCTTCCGTCAATTCCTTTAAGTTTCA		
Yeast 25S rRNA (SnR38)	[6~ FAM]ACAATGATAGGAAGAGCCGACATCGAA		
Yeast 25S rRNA (SnR24)	[6~ FAM]AACCGGATTCCCTTTCGAT		
Yeast 25S rRNA (SnR44)	[6~ FAM]TTAGCGGATTCCGACTTCCA		

Gaëlle Talhouarne

Curriculum vitae

CONTACT

E-mail: talhouarne@carnegiescience.edu

EDUCATION

2007-11	**B.S. University of Wisconsin, La Crosse, WI** Majored in Molecular Biology and Biochemistry - minor in Philosophy Graduated with highest honors
2011-pres	**Ph.D. Johns Hopkins University,** Dept. of Biology **Advisor: Dr. Joseph Gall**

RESEARCH EXPERIENCE

2008-11	Junior Research Fellow - University of Wisconsin, La Crosse Laboratory of **Dr. Scott Cooper**
2010-11	Laboratory assistant - University of Wisconsin, La Crosse Laboratory of **Dr. Sumei Liu**
2012-pres	Ph.D. candidate - Johns Hopkins University, Dept. of Biology Laboratory of **Dr. Joseph Gall**

PUBLICATIONS

Talhouarne GJS and Gall JG. **2014**. Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. *RNA* 20: 1476-1487

Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS. **2014**. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8: 1365-1379.

Liu S, Chang J, Long N, Beckwith K, Talhouarne GJ, Brooks JJ, Qu MH, Ren W, Wood JD, Cooper S, Bhargava A. **2016**. Endogenous CRF in rat large intestine mediates motor and secretory responses to stress. *Neurogastroenterol Motil.* 28:281-91

Talhouarne GJ and Gall JG. Stable lariat intronic RNAs in the cytoplasm of vertebrate cells. (***in preparation***)

Talhouarne GJ and Gall JG. 7SL RNA independent of the signal recognition particle in red blood cells. (***in preparation***)

Talhouarne GJ*, Deryusheva S* and Gall JG. Stable lariats bearing a snoRNA in vertebrate cells. (***in preparation***)

Deryusheva S*, Talhouarne GJ* and Gall JG. The missing guides. (***in preparation***)

AWARDS and FELLOWSHIPS

2008-11	Academic Initiative scholarship (x4) - University of Wisconsin, La Crosse
2009-11	Undergraduate research fellowship - University of Wisconsin, La Crosse
2010	The Dean's distinguish fellowship - University of Wisconsin, La Crosse
2011	Outstanding senior of the year in Cellular and Molecular Biology Award
2012	CMDB retreat poster award - Johns Hopkins University
2013	DuPont Teaching Award - Johns Hopkins University

INSTITUTION SERVICE

2011-14	Johns Hopkins University, student representative
2012-14	Johns Hopkins University, CMDB Program orientation organizer
2012	Organizer – Campus tour with **MInDS** (**M**entoring to **In**spired **D**iversity in **S**cience)
2013	Chairman - Johns Hopkins University, CMDB retreat
2014-15	Co-organizer – Carnegie postdoctoral and student invited speaker series
2014-16	Demo leader - **BioEyes**, promoting science to students from K-12

TEACHING

2007-10	Science tutor - **Upward Bound**, Federal program serving high school students from low-income families
2011	Lecturer – 2 Biology lecture to High school students at Baltimore Talent Development with **MindS**, JHU (**M**entoring to **In**spired **D**iversity in **S**cience)
2012	Teaching Assistant, AS.020.340, Genetics Lab, Dept. of Biology, JHU
2013	Teaching Assistant, AS.020.373, Developmental Biology, Dept. of Biology
2013	Assistant - **Center For Talented Youth**, Introduction to Bioinformatic
2016	Lecturer - Carnegie summer undergraduate lecture series

STUDENTS MENTORED

Nicole Bush (summer 2015)
Nealyn Jahangir (summers 2015 and 2016)
Yeoryios Zisopoulos (2016-pres)

TALKS and POSTERS

2010	7th Annual Celebration of Undergraduate Student Research - Poster title: "*Effects of Hibernation on Fibrinolysis and Platelet Clot Stability in 13-Lined Ground Squirrels"*
2011	**Talk** - National Conference on Undergraduate research (NCUR). Title: "*Fibrinolysis of blood clots is increased during hibernation in ground*

squirrels"

2012 International Xenopus Conference - Poster title: "*Intronic Sequences in the Cytoplasm of X. tropicalis Oocytes."*

2013 **Talk** - Dr. Gall 85th Birthday Symposium - Title: "*New discoveries in the Gall Lab*"

2013 ASCB meeting - Poster title: "*Intronic Sequences in the Cytoplasm of X. tropicalis Oocytes."*

2014 RNA Society Meeting- Poster title: *"Circular Intronic Sequences in the Cytoplasm of Xenopus Oocytes"*

2015 RNA Society Meeting - Poster title: *"Circular intronic RNA in the cytoplasm of Xenopus oocytes"*

2016 CSHL Nuclear organization & function - Poster title: *"Unexpected localization of stable intronic RNAs"*

2017 Keystone - Poster title: *"Stable lariat intronic RNA in the nucleus and the cytoplasm"*

REFERENCES

Joseph Gall, Ph.D. (Thesis advisor)
Carnegie Institution for Science
Department of Embryology
Phone: (410)-246-3017
E-mail: gall@carnegiescience.edu

Allan Spradling Ph.D.
HHMI/Carnegie Institution for Science
Department of Embryology
Phone: (410)-246-3021
E-mail: spradling@carnegiescience.edu

Sarah Woodson Ph.D.
Johns Hopkins University
Department of Biophysics
Phone: (410)-516-2015
E-mail: swoodson@jhu.edu

Karen Beemon, Ph.D.
Johns Hopkins University
Department of Biology
Phone: (410)-516-7289
E-mail: klb@jhu.edu